

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 10:11:48 ; Search time 2486.45 seconds
(without alignments)
18113.913 Million cell updates/sec

Title: US-09-667-130-1

Perfect score: 3337

Sequence: 1 gaattccggtaaagtaacaa.....tcttataaaataataattc 3337

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_esti:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	435.8	13.1	633	12	AZ569608 263PVD09
2	220	6.6	475	10	BM029212 IpSkp0005
3	214.2	6.4	691	12	BH040136 RPCI-24-2
4	209	6.3	908	12	AZ548467 ENTER30TR
5	205	6.1	843	12	AZ551618 ENT5V54TR
6	203.2	6.1	877	12	AZ531291 ENTBQ34TR
7	200.6	6.0	890	12	BH146886 ENTPK48TR
8	200.6	6.0	912	12	AZ551092 ENTJ22TR
9	199.8	6.0	931	12	BH160272 ENTQV49TR
10	197	5.9	605	12	AZ640388 IM0502E21
11	195.4	5.9	880	12	AZ529191 ENTBV68TR
12	193.8	5.8	735	12	CNS04NSM
13	192.4	5.8	906	12	BH153606 ENTPS83TF
14	192	5.8	501	12	FR0048173
15	191.8	5.7	778	10	BM170084
16	190.4	5.7	905	12	AZ550256 ENTEV58TR
17	189.4	5.7	443	10	BM374465 EBP103_SQ

C 18	186.8	5.6	849	12	AZ546009
19	185.8	5.6	715	10	BM170799
20	184.2	5.5	650	12	AZ337339
21	182.2	5.5	421	12	AZ113646
22	181.2	5.4	596	12	AZ460541
23	180.4	5.4	494	12	FR0048073
24	180.4	5.4	816	12	AZ535744
25	178.4	5.3	976	12	BH149983
26	178	5.3	891	12	AZ683582
27	176	5.3	942	12	BH148582
28	175.8	5.3	879	12	AZ550718
29	175.2	5.3	832	10	BM170146
30	173.6	5.2	718	12	AZ972907
31	172.8	5.2	540	12	AZ813205
32	171.2	5.1	669	12	AZ392101
33	171.2	5.1	898	12	CNS04ALV
34	170.4	5.1	773	12	CNS01VTG
35	168.4	5.0	824	12	AZ185454
36	167.4	5.0	481	10	BM441017
37	167.2	5.0	677	10	BM161774
38	166.8	5.0	619	12	FR0047601
39	166.8	5.0	709	10	BM165622
40	166.4	5.0	617	12	AZ296172
41	166.2	5.0	332	10	BE437454
42	166	5.0	774	12	AZ658872
43	165.4	5.0	746	10	BM165802
44	165	4.9	938	12	CNS07CG2
45	164.6	4.9	868	12	BH161951

ALIGNMENTS

RESULT 1

AZ569608 263PVD09 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
LOCUS 633 bp DNA linear GSS 15-MAY-2001
DEFINITION Plasmodium vivax
ACCESSION AZ569608
VERSION AZ569608.1 GI:13979872
KEYWORDS GSS.
SOURCE Malaria parasite P. vivax.
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 633)
AUTHORS Carlton,J.M.-R. and Dame,J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: Shotgun.

FEATURES

Location/Qualifiers
1..633
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"
/db_xref="taxon:5855"
/clone_lib="Pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Saimiri boliviensis"
/note="vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmodipur filter, followed by passage through a column of pre-wet

Whatman CFl1 powder (1:2 ratio volume of blood to CFl1), and finally centrifuged through a 50% Percoll density cushion. Purified DNA was digested with mung bean nuclease in the presence of 44% formamide at 50°C as described (Vernick, K.D., Imberski, R.B., and McCutchan, T.F. 1988. Nucleic Acids Research 16:6883-6896). Digested DNA was blunt-ended using T4 DNA Polymerase and size fractionated over a Sepharose CL-2B column. Fractions in the size range 500bp-4kb were ligated into the Eco RV site of pBluescript SK(+), and E. coli XL-10 Gold transformed with the ligation mixture."

```

BASE COUNT      206 a  84 c  196 g  146 t      1 others
ORIGIN

Query Match      13.1%; Score 435.8; DB 12; Length 633;
Best Local Similarity 99.3%; Pred. No. 5.7e-47;
Matches 448; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

Qy 261 gatgttagaggtggaggatccttcaaacacgctgttagaattagaagaggaatttt 320
Dy 1 GATGTTGTAGAGTGGAGGATCCTTCAACACGCGTTTGAATTTAGAAGAGGAAAAATTT 60

Qy 321 gatgagaattcagtgatgataactcttttagatgctactccccggaagatgactttgcc 380
Dy 61 GATGAGAATTCAGGTGATGATGAACCTCTTTTAGATGCTACCCCGGAGATGACTTTGCC 120

Qy 381 ttaacagatttgccaattgaagcagatgaggaagtcacacgaacgttagatggagtgaa 440
Dy 121 TTACAGATTTGCCAATTGAAGACGATGAGGAAGTCAACGAAACGTTAGATGGAGGTGAA 180

Qy 441 tcattagagaggtttccactgaagatatgaaacagaagatggctcaacagatgatacg 500
Dy 181 TCATTAGAGAGGTTTCCACTGAAGATATGAAACAGAAAGATGCTCAACAGATGATACG 240

Qy 501 gaaacagaagaagactacctgtgatatgaaagagaagaagctggcgatagaa 560
Dy 241 GAAACAGAAGAAGGACTACCTGTGTATGGAAGGAGAAGAAGTGGCGATATGGA 300

Qy 561 gcagggaagaagctggtgatttgaagcagggggaagaaactgctgatttgaagcagg 620
Dy 301 GCAGGGAAGAAGCTGTTGATTTGGAGCAGGGAAGAACTGCGGATTTGGAAGCAGGG 360

Qy 621 gaagaactggcgatttggagcagggggaagaagctggttatttggagcagggggaaga 680
Dy 361 GAAGAAGCTGGCGATTTGGAAGCAGGGAAGAGCTGCTGATTGGAAGCAGGGGAAGA 420

Qy 681 actgacatttggagcagggggaagaactg 711
Dy 421 ACTTGGCATTTGG-AGCATGGGAAGAACTG 450

```

```

RESULT 2
BM029212      475 bp  mRNA  linear  EST 05-NOV-2001
LOCUS      ipSkn00051 Skin cDNA library Ictalurus punctatus cDNA 5', mRNA
DEFINITION      sequence.
ACCESSION      BM029212
VERSION      BM029212.1 GI:16742782
KEYWORDS      EST.
SOURCE      channel catfish.
ORGANISM      Ictalurus punctatus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
Ictaluridae; Ictalurus.
1 (bases 1 to 475)
Karsi,A., Cao,D., Li,P., Ju,Z., Kocabas,A., Feng,J., Patterson,A.,
Mickett,K.D. and Liu,Z.
Transcriptome analysis of channel catfish (Ictalurus punctatus):
Initial analysis of gene expression and microsatellite-containing
cDNAs in the skin
Unpublished (2001)
Contact: Liu ZJ

```

The Fish Molecular Genetics and Biotechnology Laboratory,
Department of Fisheries and Allied Aquacultures and Program of Cell
and Molecular Biosciences
Auburn University
203 Swingle Hall, Auburn University, Auburn, AL 36849, USA
Tel: 334 844 4054
Fax: 334 844 9208
Email: zliu@acesag.auburn.edu
Seq primer: M13 Reverse.

```

FEATURES
    source      1..475
                /organism="Ictalurus punctatus"
                /db_xref="taxon:7998"
                /clone_lib="Skin cDNA library"
                /note="Organ: Skin; Vector: pSport1; Site_1: NotI; Site_2:
                Sall"
BASE COUNT      229 a  77 c  154 g  15 t
ORIGIN

```

```

Query Match      6.6%; Score 220; DB 10; Length 475;
Best Local Similarity 69.0%; Pred. No. 3.2e-19;
Matches 301; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

Qy 2019 gaattagaagaaccacagggagaagtagaagaacccgtagagggcgagaactgca 2078
Dy 1 GAAGAAGAAGAAGCAGCAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCAGCAGCA 60

Qy 2079 gaaggagaagaagtgggaaggttaccctgcagaagtagaagaagtggagaagtagcctgca 2138
Dy 61 GCAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 120

Qy 2139 gaagtgaagaagtgggaaggttaccagagaagtagaagaggtaccgccgaagtagaa 2198
Dy 121 GAACCAGAAGAAGAAGCAGCAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGCAGCA 180

Qy 2199 gaagtgaagaaggttaccagaagaagtgggaaggttaccagaagaagtgggaaggtacca 2258
Dy 181 GCAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 240

Qy 2259 gaagaagtgaagaaggttaccagaagaagtgggaaggtggagaagtagaagaagtagag 2318
Dy 241 GAAGAAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 300

Qy 2319 gtaccagcggtagtagaagttagaagttaccagcggtagtagaagaagaggtgtccagaaga 2378
Dy 301 GCAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 360

Qy 2379 gtagaagaagaagaagaaggaaggaaccagtagagaagaagatgtattacaattagta 2438
Dy 361 GCAGCAGAAGAAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA 420

Qy 2439 ataccatcggaagaag 2454
Dy 421 GCAGCAGCAGAAGAAG 436

```

```

RESULT 3
BM040136/c
LOCUS      RPCI-24-255A23.TV RPCI-24 Mus musculus genomic clone RPCI-24-255A23
DEFINITION      , DNA sequence.
ACCESSION      BM040136
VERSION      BM040136.1 GI:14818806
KEYWORDS      GSS.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 691)
Zhao,S., Niernan,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Kroi,M., Shvartsbeyn,A., Gebregorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.

```

Qy 1915 tagaaqgaaggaqaaccaacggaagaagaagccaaqagaaggagagcacaagaaggag 1974

Db	131	AAGATGATGAAGATGAAGATGATGAAGACGATGAAGACGACGAATGATGAATTGAAGATG	190
Qy	1975	aagtgcagaagaagaattagagcaactccagaggacgatttgcgaattagaagaaccaa	2034
Db	191	ATGATGATGAAGAAGAAGAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAAT	250
Qy	2035	caggagaagaagttagaagaacccttagaggcgaagaactcgagaagagaagaagtgg	2094
Db	251	ATGAATTTAGAAGATGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGAAGATG	310
Qy	2095	aagagggtacctcgcaagaatagaagaagtggagaaggtacctcgagaagttagaagaagtgg	2154
Db	311	ATGATGAAGATGATGATGAACAAGAAGATGAAGAAGATGATGAACAACATGATGAAGATG	370
Qy	2155	aagagggtaccagaagaagttagaagaaggttaccgcgagaagtagagaagaatggaagaaggtac	2214
Db	371	ATGATGAAGAAGAAGAAGAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAATATG	430
Qy	2215	cagaagaagtggagaaggtaccagaagaagtggagaaggttaccagaagaagtggagaaggtg	2274
Db	431	AATTGAAGATGATGATGAACAAGAAGAAGATGATGATGAACAAGAAGAAGATGATGATG	490
Qy	2275	taccagaagaagtggagaagaatggaagaagttagaagaagtagagaaggtaccagcggtagtag	2334
Db	491	AAGATGATGATGAAGAAGAAGAAGATGAAGATGATGATGAAGAAGAAGAAGATGAAGATGATG	550
Qy	2335	aagttagaagttaccagcggtagtagaagaagaaggtggccagaagaagttagaagaagaagaag	2394
Db	551	AAGAAGAAGATGATGATGAACAAGAAGATGAAGAAGAAGAAGATGATCAAGAAGATGATG	610
Qy	2395	aagagggaaccagtagaggaagaagatgtattacaattagtaataccatcggaagaag	2454
Db	611	AAGAAGAAGAAGAAGATGATGAAGACGACGAATATGAATTTAGAAGATGATGAAGAAGATG	670
Qy	2455	atatacaattagacaaaccaaagaagacgaattaggctctggaaattttatctatctatcgcg	2514
Db	671	ATGAAGAAGATGATGAACAAGAAGATGATGAAGAAGAAGAAGATGATCATGATGATGATG	730
Qy	2515	acatgcactaccagaagttccaaaggaattattggaagaagaagaagaactgcagtgct	2574
Db	731	AANGAAGTGTGTCGAAGATGATGAAGAAGAAGATGAAGATGATGAAGATGATGAAGAAGATG	790
Qy	2575	atccattgaaaccagaagaattttgaaagggaagattccaaactctacaga	2623
Db	791	AAGATGATGAAGATGATCAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGA	839

RESULT	5
AZ551618/c	
LOCUS	AZ551618.1 linear GSS 14-NOV-2000
DEFINITION	ENTD54TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION	AZ551618
VERSION	AZ551618
KEYWORDS	GSS.
SOURCE	AZ551618.1 GI:11176919
ORGANISM	Entamoeba histolytica. Eukaryota; Entamoebidae; Entamoeba.
REFERENCE	1 (bases 1 to 843)
AUTHORS	Lofthus,B., Van Aken,S. and Fraser,C.
TITLE	Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library
JOURNAL	Unpublished (2000)
COMMENT	Contact: Brendan J Lofthus Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0208 Fax: 301 838 3543 Email: bjloftus@igr.org Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer: M13-Reverse					
Class: Shotgun					
High quality sequence start: 39					
High quality sequence stop: 838.					
FEATURES					
source					
1..843					
/organism="Entamoeba histolytica"					
/strain="HMI:IMSS"					
/db_xref="taxon:5759"					
/note="Vector: pHOsi; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD."					
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."					
BASE COUNT 155 a 266 c 18 g 404 t					
ORIGIN					
Query Match 6.1%; Score 205; DB 12; Length 843;					
Best Local Similarity 55.6%; Pred. No. 2.3e-17;					
Matches 394; Conservative 0; Mismatches 315; Indels 0; Gaps 0;					
Qy	1855	aagaggcagatttcgaattagaagaaactgcagaagctccagaggaagaggaatagtat	1914		
Db	807	ATGATGTGAAGACAGATCATGATCAGAAAGATGATGAACAGATGATGAAGAAGAG	748		
Qy	1915	tagaagagaagggagaaccaaaggaaagagcgcaaggaaggagaccacaagaagag	1974		
Db	747	AGTATGATGAAGATCAAGATGATGAAGACGATGAAGACGACGAATATCATTTAAGATG	688		
Qy	1975	aagtgccagaagaagaattagagccaactccagaggacgattccgaattagaagaacca	2034		
Db	687	ATGATGTGAAGAAGAACAGATGATGATGAAGATGATGAAGACGNTGAGCCGCCAAT	628		
Qy	2035	caggagaagaagttagaagaaaccttagagggcgcaagaaactgcgaagaggagaagaagtg	2094		
Db	627	ATGAATTAGAAGATGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGAATG	568		
Qy	2095	aagaggctacctgcagaagtagaagaagtggaaagggtacctgcgaagtagaagaagtgg	2154		
Db	567	ATGATGAAGATGATGATGAAGAAGAAGATGAAGTAAGATGATGAAGAAGATGATGAAGATG	508		
Qy	2155	aagaggctaccagaagaagttagaagaggttaccocgcagaagtagaagaagtggaagggta	2214		
Db	507	ATGATGAAGAAGAAGAGATGATGATGAAGATGATGAAGACGATGAAGACGACCATAATG	448		
Qy	2215	cagaagaagtggaaaggttaccagaagaagtggaaaggggtaccagaagaagtggaagagg	2274		
Db	447	AATTAGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGAAGATGATGATGATG	388		
Qy	2275	taccagaagaagtggaaagaagttagaagaagttagaagtagaggttaccagcggtagtag	2334		
Db	387	ANGATGATGTGAAGAGAAGACATGAGATGATGATGAAGAGAAGATGAAGATGATGATG	328		
Qy	2335	aagttagaagttaccagcggttagtagaagaagaaggtgcccagaagaagtagaagaagaaga	2394		
Db	327	AAGAAGAAGATGATGATGAAGAAGAAGATGAAGAAGAAGAAGATGATGAAGAAGATGATG	268		
Qy	2395	aagggagaaccagtagagagaagaacatgtattacaattagtaataccatcggaagaag	2454		
Db	267	AAGAAGAAGAAGATGATGATGAACGACGCAATATGAATTAGAAGATGATGAAGAAGATG	208		
Qy	2455	atatcacattagacaaaccaaagaacacgataggctcgtggaattttatctcatcgcg	2514		


```
Db 207 ATGAAGAAGATGATGAAGAAGAGATGATGAAGAAGAGATGATGATGATGATG 148
QY 2515 acatgcactaccagagcttccaaaggaatttatggaagaagaagaaga 2563
Db 147 AACAGATGATGATGAAGATGATGAAGAAGATGATGAAGAAGA 99

RESULT 6
AZ531291/c 877 bp DNA linear GSS 03-NOV-2000
LOCUS ENTBQ34TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION AZ531291
VERSION AZ531291.1 GI:11085838
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 877)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 22
High quality sequence stop: 829.
Location/Qualifiers
1..877
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT

156 a 268 c 29 g 424 t

ORIGIN

```
Query Match 6.1%; Score 203.2; DB 12; Length 877;
Best Local Similarity 54.9%; Pred. No. 3.8e-17;
Matches 400; Conservative 0; Mismatches 328; Indels 0; Gaps 0;

QY 1749 gatgcaggagggctgtaacagttaccataaaggagcaccctgtacaagtaccaggga 1808
Db 815 GAAGATGAAGACCAATATGATTAAGAGAGAGAGAGATGATGATGATGAAGAAGAA 756
QY 1809 gtagggccgcgcgaagaagtgcgaacgcggaattgatgcacactccaaggagcgatttc 1868
Db 755 GAAGATGATGAAGATCATGATGATGAAGAAGATGATGATGAAGAAGATGATGATGAAGAA 696
QY 1869 gaattagaagggaactcagaagctccagagggaagggaattagattagaaggagaaga 1928
```

```
Db 695 GATGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGAGATGATGAAGAT 636
QY 1929 gaaccaacggaagaagaagccaaagagagagagagagagagagagagagagagagagaa 1988
Db 635 GAAGATGATGAAGACCGATGAAGACGACGCAATATGATTAAGATGATGATGAAGAA 576
QY 1989 gaattagaggcaactccagaggacgatttcgaatttagaagaacacacaggaagaagta 2048
Db 575 GAAGAAGATGATGAAGATGATGAAGACGATGAAGACGACGACGCAATATGAATTAGAAGAT 516
QY 2049 gaagaacacctagaggcgcaagaagaactgcgaaggaggaagaagtgcgaagggtacctcgca 2108
Db 515 GATGATGATGATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGATGAAGATGAT 456
QY 2109 gaagtgaagaagtgggaagggttacctgcgaagtagaagaagtgggaagggtaccagaa 2168
Db 455 GATGAAGAAGACATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGAA 396
QY 2169 gaagtgaagaagggtaccgcgcagagtagaagaagtggaaggtaccagagaagtgga 2228
Db 395 GAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGCAATATGAATTAGAAGATGAT 336
QY 2229 gaggtaccagaagaagtgggaagggtaccagagaagaagtgcgaagggtaccagagaagtgc 2288
Db 335 GATGAAGAAGAAGATGATGATGAAGAAGAAGATGATGATGAAGAAGATGATGAAGAAGATGAT 276
QY 2289 gaagaagtgggaagaagttagaagaagtagaggtaccagcggttagtagaagtagaagtagacca 2348
Db 275 GAAGAAGATGAAGATGATGATGAAGAAGAAGATGAAGATGAAGATGAAGATGAAGATGAT 216
QY 2349 gcggttagtagaagaagaaggtgcccagaagaagtagaagaagaagaagaagaagaagtagacca 2408
Db 215 GATGAAGAAGAAGATGAAGAAGAAGAGATGATGAAGAAGATGATGAAGAAGAAGAGAT 156
QY 2409 gttagagaagaagatgtattacattagtaattaccatccgcgaagaagatatcaattagac 2468
Db 155 GATGATGAAGACCGACGAATATGAATTAGAAGATGATGAAGAAGATGATGAAGAAGATGAT 96
QY 2469 aaacacaaa 2476
Db 95 GAAGAAGA 88
```

RESULT 7

```
BH146886
LOCUS ENTBQ48TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, DNA sequence.
ACCESSION BH146886
VERSION BH146886.1 GI:15302963
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 890)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HM1:IMSS sheared DNA library (2001)
COMMENT Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 880.
Location/Qualifiers
```

FEATURES


```

RESULT 10
AZ640388/c 605 bp DNA linear GSS 14-DEC-2000
LOCUS
DEFINITION
LM0502E21F Mouse 10kb plasmid UUGCLM library Mus musculus genomic
clone UUGCLM0502E21 F, DNA sequence.
ACCESSION
AZ640388
VERSION
AZ640388.1 GI:11763788
KEYWORDS
GSS.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (Bases 1 to 605)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
, M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0502 row: E column: 21
Seq primer: CGTGTAAACGACGCGCAGT
Class: plasmid ends
High quality sequence stop: 605.
Location/Qualifiers
1..605
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGCLM0502E21"
/clone_lib="Mouse 10kb plasmid UUGCLM library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of PWD42 (g114732114[gblAF129072.1]), a copy-number
inducible derivative of plasmid RL. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
13 a 253 c 136 g 203 t
BASE COUNT
ORIGIN

Query Match 5.98; Score 197; DB 12; Length 605;
Best Local Similarity 59.38; Pred. No. 2.7e-16;
Matches 335; Conservative 0; Mismatches 230; Indels 0; Gaps 0;

Qy 1874 agaaggaactcagaagtcacaggaagaggaattatgattatagaagagaggaacc 1933
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 570 AGCAGGACGACGTAGNAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 511

```

```

Qy 1934 aacggaagaagagccaagagagagagccaacagaagaagtgccagagaagaatt 1993
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 510 AGCAGGACGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGG 451
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 1994 agaggaactcagagagcagatttcgaattagaagaacacacagagagaagtagaaga 2053
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 450 AGCAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGG 391
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2054 aaccgtagaggcggaagaactgcagaagagagaagaagtgaagaggttacctgcagaagt 2113
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 390 AGCAGCAGCAGCAGCAGCAGGCGGCGAGCAGGCGGCGAGCAGCAGAAAGCAGGAGCAGCAGG 331
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2114 agaagaagtggaaagggttacctgcagaagtagaagaagtggaaagggttaccagaagaagt 2173
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 330 AGCAGCAGGCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGCGAGCAGCAGCAGCAG 271
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2174 agaagaggttacccgcagaagtagaagaagtggaaagggttaccagaagaagtggaaagagt 2233
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 270 AGCAGCAGGCGGCGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGCGGCGAGCAGCAGCAGT 211
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2234 accagaagaagtggaaagggttaccagaagaagtggaaagggttaccagaagaagtggaaaga 2293
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 AGCAGTAGGACGACGTAGGACGAGCAGGAGCAGCAGCAGCAGCAGCAGCAGGCGGCGAGC 151
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2294 agtggaaagaagtagaagaagttagagggttaccagcggttagtagaagttagatccagcggt 2353
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 150 AGGAGCAGCAGAGAAGCAGCAGGCGGCGAGCAGCAGCAGCAGCAGCAGCAGGCGGCGAGCAGC 91
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2354 agtagaagaagaggtgccagaagaagtagaagaagaagaagaagaaggaaccagtagaga 2413
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 90 AGCAGAGCAGGAGCAGTAGGAGCAGTAGGAGCAGCAGCAGCAGCAGCAGCAGGAGCAGCAGGAGC 31
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 2414 ggaagaagatgtattacaattagta 2438
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 30 AGCAGGACGACGTAGGAGCAGGTAGGA 6
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AZ529191/c 880 bp DNA linear GSS 03-NOV-2000
LOCUS
DEFINITION
ENTBV68TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION
AZ529191
VERSION
AZ529191.1 GI:11081835
KEYWORDS
GSS.
SOURCE
Entamoeba histolytica.
ORGANISM
Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
1 (Bases 1 to 880)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlloftus@tigr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 673.
Location/Qualifiers
1..880
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"

FEATURES
source

```


Qy	2328	gtagtagaagtgaaggtaccacgcgdtagtgaagaagagggtgccagaagaagtagaagaa	2387
Db	423	GCAGTGTAGTAGCAGTAGTAGTAGTAGTAGTAGTAGTAGTAGTGACGCGMAATWGTA	482
Qy	2388	gaagaagaagggaagcaaccagtgaggaagaagtgtattacaattagtataaccatcg	2447
Db	483	GCAGTGTAGTAGYAGCTGCAGTAGTAGCACGTAGTTAGTTTTAGCAGTAGTAGTAGTAGCA	542
Qy	2448	gaagaagatatacaattag	2466
Db	543	GTAGTTGTTGTATTAGTAG	561

RESULT 13
BH153606/c
LOCUS
BH153606
906 bp
DNA
linear
GSS 24-SEP-2001
DEFINITION
Entamoeba histolytica Sheared DNA
Entamoeba histolytica
genomic, DNA sequence.

Accession
Version
Keywords
Source
Organism
Reference
Authors
Title

Unpublished (2001)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org
Clones are derived from the *Entamoeba histolytica* HML:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 733.

```

FEATURES
source
1. 906
Location/Qualifiers
/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

Query Match	5.8%	Score 192.4;	DB 12;	Length 906;
Best Local Similarity	55.1%	Pred. No. 9.2e-16;		
Matches 376;	Conservative	0;	Mismatches 306;	Indels 0;
			Gaps	0;

Qy 1743 gatcctgatgcaggagaggctgaacagtaccatcaaaagaagcacctgtacaagtacca 1802
||| ||||| | | | | | | | | | | | |
Db 733 GATGATGATGAAGATGACGATGATGATGAAGAAGATGATGATGCCGATGAGAAAGACGAT 674
||| ||||| | | | | | | | | | | | |

Qy	1803	gtggcagtaggccccgcgcaagaagtgc	caacggaagattgtgcaactccaagagac	1862
Db	673	GAAGAACACGATGATGAAGACCGATGATGAAGACCGATGATGAAGACCGATGATGAAGACCGAT	614	
Qy	1863	gatttcgaattagaagaactcagaagctccagaggagagaaattagattagaagga	1922	
Db	613	GAAGAGCGCGATGATGAAGACGATGATGAAGAGACGATGATGAATTTGTAATTGAAGAC	554	
Qy	1923	gaaggagaacccaacggaagaagc	ccaagagagagagccaaacagaaggaagtgcca	1982
Db	553	GATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACG	494	
Qy	1983	gaagaagaattagaggccaactccagaggagcattctgaaattagaagaacccaacggagaa	2042	
Db	493	GATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACG	434	
Qy	2043	gaagtagaagaacccgtagagggcgaagaaacctgcagaaagagagaagtgaagaggtta	2102	
Db	433	GATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACG	374	
Qy	2103	ccctgcagaagttagaagaagtggaaaggttacctgcagaagtagaagaagtggaaagaggtta	2162	
Db	373	GATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACG	314	
Qy	2163	ccagaagaagttagaagaaggtaccgcgagaagtagaagaagtggaaaggtaccagaagaa	2222	
Db	313	GAAGATGATGAAGATGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACCGATGATGAAGAACG	254	
Qy	2223	gtggaagaggtaccagaagaagtggaaaggtaccagaagaagtggaaaggtaccaggaa	2282	
Db	253	GAAGACGCGATGAACACCGACGATGAAGACCGATGAAGACCGATGAAGACCGATGAAGACCGAT	194	
Qy	2283	gaagtggaagaagtggaagaagttagaagaagttagagggtaccagcggtagtagaagtagaa	2342	
Db	193	GAAGAAGACCGATGAAGAAGACCGATGAAGAAGACCGATGAAGAAGACCGATGAAGAAGACCGAT	134	
Qy	2343	gtaccagcggtagtagaagaaggtgc	ccaagaagtagaagaagaaggaagagaa	2402
Db	133	GAGAGACCGATGAAGAAGACCGATGAAGAACCGATGAAGAACCGATGAAGAACCGATGAAGAACCGAT	74	
Qy	2403	gaaccagtagaagaagaagtgc	2424	
Db	73	GAAGAAGACCGATGAAGAAGACG	52	

RESULT	14
LOCUS	FR0048173/c
DEFINITION	FR0048173 Fugu rubripes GSS sequence, clone 264E22ca9, genomic survey sequence.
ACCESSION	AL444958
VERSION	AL444958.1 GI:12053458
KEYWORDS	GSS; genome survey sequence.
SOURCE	Takifugu rubripes.
ORGANISM	Takifugu rubripes ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 Tetraodonlineage; hakiungu.
1 (bases 1 to 501)
Clark, M.S.
Direct Submission
Submitted (29-SEP-2000) MRC Human Genome Mapping Project Resource
Centre Hinxton, Cambridge, CB10 1SB. UK Email:
biohelp@hgmrc.ac.uk
Vector: pBluescript II KS
V-type: phagemid
PRIMER: KS

FEATURES

One pass dye-terminator sequencing of BAC (pBelobACII) cloned genomic sequence

The BACs can be obtained from <http://www.incyte.com>.
Location/Qualifiers

FEATURES		Location/Qualifiers		JOURNAL		COMMENT						
Source		1. .443		Unpublished (2000)		Contact: Brendan J Loftus						
		/organism="Hordeum vulgare"				Department of Eukaryotic Genomics						
		/cultivar="Optic"				The Institute for Genomic Research						
		/db_xref="taxon:4513"				9712 Medical Center Dr., Rockville, MD 20850, USA						
		/clone="EBp103_SQ003_L06"				Tel: 301 838 0208						
		/clone_lib="IGF Barley EBp103 library"				Fax: 301 838 3543						
		/tissue_type="Pistils"				Email: bjloftus@tigr.org						
		/dev_stage="4 days post anthesis"				Clones are derived from the Entamoeba histolytica HM1:IMSS sheared						
		/lab_host="DH10B"				DNA library						
		/note="Vector: pSPORT1; Site_1: Sal I; Site_2: Not I;				Seq primer: M13-Forward						
		Non-normalised library, directionally cloned into pSPORT1.				Class: shotgun						
		Derived from pistils dissected from developing grains (4				High quality sequence start: 26						
		days post anthesis) in glasshouse grown barley plants.				High quality sequence stop: 796.						
		Developed as part of the barley transcriptome resources of										
		BBSRC/SEERAD funded cereal IGF (Investigating Gene										
		Function) project."										
BASE COUNT		287 a	8 c 126 g	22 t								
ORIGIN												
Query Match		5.7%; Score 189.4; DB 10; Length 443;										
Best Local Similarity		66.2%; Pred. No. 2.8e-15;										
Matches	288;	Conservative	0;	Mismatches	146;	Indels	1; Gaps 1;					
QY	2042	agaagtgaagaaccgtagaggcggaagaactgcagaaggagaagaagtgaaggt	2101									
Db	9	AGAAGTAGAAGAAGAGGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA	68									
QY	2102	acctgcagaagtgaagaagtgaagaaggtacctgcagaagtgaagaaggtgaaggt	2161									
Db	69	AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA	128									
QY	2162	accagaagaagtgaagaaggtaccgcagaagtgcagaaggtgaagaaggtaccaga	2221									
Db	129	AGAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA	188									
QY	2222	agtggaagaggtaccagaagaagtgaagaaggtaccagaagaagtgaaggtaccaga	2281									
Db	189	AGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA	248									
QY	2282	agaagtgaagaagtgaagaagtgaagaaggtaccgcaggtaccgcaggttagtaga	2341									
Db	249	AGAAGCACACAGACGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA	307									
QY	2342	agtaccagcggtagtagaagaaggtgccagaagtagagaagaagaagaagga	2401									
Db	308	ATAACATGAGGATATATATAGAGAAGAAGAATAGAGACGAAGAAGAATA	367									
QY	2402	agaaccagtagagaagaagtgattaccattagtaataccatcggaagaagatacata	2461									
Db	368	AAAAGAAAAGAATAAGAATAAGGATGAGATAAGAAGCACAGACGAAGAAGA	427									
QY	2462	attagacaacccaaa	2476									
Db	428	AGAAGAAGAAGAAGA	442									
RESULT 18												
AZ546009/C		849 bp		DNA		linear						
LOCUS		GSS 14-NOV-2000										
DEFINITION		ENTFW537F Entamoeba histolytica Sheared DNA Entamoeba histolytica										
		genomic, DNA sequence.										
ACCESSION		AZ546009										
VERSION	AZ546009.1	GI:11167130										
KEYWORDS		GSS.										
SOURCE		Entamoeba histolytica.										
ORGANISM		Entamoeba histolytica										
REFERENCE		1 (bases 1 to 849)										
AUTHORS		Loftus,B., Van Aken,S. and Fraser,C.										
TITLE		Determination of clone end sequences from Entamoeba histolytica										
		HM1:IMSS sheared DNA library										

Unpublished (2000)		Contact: Brendan J Loftus		Department of Eukaryotic Genomics		The Institute for Genomic Research		9712 Medical Center Dr., Rockville, MD 20850, USA		Tel: 301 838 0208		Fax: 301 838 3543		Email: bjloftus@tigr.org		Clones are derived from the Entamoeba histolytica HM1:IMSS sheared		DNA library		Seq primer: M13-Forward		Class: shotgun		High quality sequence start: 26		High quality sequence stop: 796.		Location/Qualifiers		1. .849		/organism="Entamoeba histolytica"		/strain="HM1:IMSS"		/db_xref="taxon:5759"		/clone_lib="Entamoeba histolytica Sheared DNA"		/note="Vector: pHOSt1; Site_1: Bst I; Constructed at The		Institute for Genomic Research (TIGR), Rockville, MD.		Genomic DNA isolated from broth cultures of E. histolyt		using a method described by Clark and Diamond (Clark,		C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a		method for isolate identification. Exp. Parasitol.		77:450.). The DNA was mechanically sheared to give a		tight size distribution (~2 kb). The v + i method used for		the library construction is described in detail in Smith,		H.O. and Venter, J.C. (Making small insert libraries for		whole genome shotgun sequencing projects. In Genome		Sequencing: A Practical Approach, eds. M. Vaudin and B.		Barell, Oxford University Press, 1999)."		BASE COUNT		199 a	236 c	43 g	371 t	ORIGIN		Query Match		5.6%; Score 186.8; DB 12; Length 849;		Best Local Similarity		53.8%; Pred. NO. 4.9e-15;		Matches		386;	Conservative	0;	Mismatches	332;	Indels	0;	Gaps	0;																																																																																																																																																																																																																																																																																																																																																																																																																																																																																	
QY	1846	tgcaactccaagaggacgatttcgaattagaaggaactgcagaagctccagaggaaggag	1905																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																										

JOURNAL
COMMENT

Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SIC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0068 row: 8 column: 03
Seq primer: CGTGTAAACGACGGCCAGT
Class: plasmid ends
High quality sequence stop: 650.
Location/Qualifiers
1. .650
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0068B03"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, TI-resistant, P-"
/note="Vector: pWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (g114732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

FEATURES
source

17 a 306 c 107 g 220 t
BASE COUNT
ORIGIN
Query Match 5.5%; Score 184.2; DB 12; Length 650;
Best Local Similarity 57.3%; Pred. No. 1.2e-14;
Matches 333; Conservative 0; Mismatches 248; Indels 0; Gaps 0;
Qy 1874 agaaggaactgcagaagctccagaggaaggaattagattagagaaggaaggaacc 1933
Db 649 AGCAGGAGGAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 590
Qy 1934 aacggaaggaagcagaaggaaggaagcacaaggaaggaaggaaggaaggaatt 1993
Db 589 AGGAGCAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 530
Qy 1994 agaggcaactccagagacatttcgaattagaagaacacaggaaggaaggaagga 2053
Db 529 AGCAGGAGGAGGAGCAGCAGTAGCAGTAGCAGTAGCAGGAGGAGGAGGAGGAGGAGG 470
Qy 2054 aaccgtagagggcgaagaactgcagaaggaaggaaggaaggaaggaaggaaggaatt 2113
Db 469 AGGAGGAGCAGGAGCAGCAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 410
Qy 2114 agaagaagtggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2173
Db 409 AGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 350
Qy 2174 agaagaaggtaccgcagaaggaaggaaggaaggaaggaaggaaggaaggaagga 2233
Db 349 AGCAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 290

QY 2234 accagaagaagtgaagaggtaccagaagaagtgaagaggtaccagaagaagtgaaga 2293
Db 289 AGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 230
QY 2294 agtggagaagtagaagaagtagaggtaccagcggttagtgaagtagaagtagaccagcggt 2353
Db 229 AGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 170
QY 2354 agtagaagaagaagtgccagaagaagtagaagaagaagaagaagaagaaccagtaga 2413
Db 169 AGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGT 110
QY 2414 ggaagaagatgtattacaattagtaataaccatcggaagaag 2454
Db 109 AGGAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGCAG 69

RESULT 21
AZ113646/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AZ113646 421 bp DNA linear GSS 12-MAY-2000
RPCI-23-447D2.TJ RPCI-23 Mus musculus genomic clone RPCI-23-447D2
DNA sequence.
AZ113646
AZ113646.1 GI:7773245
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 421)
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Aklnret
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other_GSSs: RPCI-23-447D2.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 447 row: D column: 2
Seq primer: SP6
Class: BAC ends.

FEATURES
source

Location/Qualifiers
1. .421
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-447D2"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site:1;
ECORI; Site:2: EcorI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcorI and EcorI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
15 a 134 c 24 g 248 t
BASE COUNT
ORIGIN

[illegible]

RESULT	25	
LOCUS	BH149983	
DEFINITION	BH149983	976 bp DNA linear GSS 27-AUG-2001
ACCSSION	ENTQD93TF	Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
VERSION	BH149983	
KEYWORDS	BH149983..1	GI:15311185
SOURCE	GSS.	
ORGANISM	Entamoeba histolytica.	
REFERENCE	Entamoeba histolytica	
AUTHORS	Eukaryota; Entamoebidae; Entamoeba.	
TITLE	1 (bases 1 to 976)	
JOURNAL	Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.	
COMMENT	Determination of clone end sequences from Entamoeba histolytica	
	HM1:TMSS sheared DNA library (2001)	
	Unpublished (2001)	
	Contact: Brendan J Loftus	

Email: bjoitusteig1@rny.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 38
 High quality sequence stop: 740.
 Location/Qualifiers
 1. .976
 /organism="Entamoeba histolytica"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica sheared DNA"
 /note="Vector: pHOSt; Site_1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome
 Sequencing: A Practical Approach, eds. M. Vaudin and B.

	Barell, Oxford University Press, 1999).	"
BASE COUNT ORIGIN	415 a 63 c 280 g 218 t	
Query Match	5.3%; Score 178.4; DB 12; Length 976;	
Best Local Similarity	54.6%; Pred. No. 5.7e-14;	
Matches 356; Conservative	0; Mismatches 296; Indels	0; Gaps 0;
Qy 1837	aagaattgatgcaactccaagagacgatttcgaattagaaggaaactcgagaactccag	18956
Db 155	ATGAAGATTTCGATTTTAATGATGATGAGTATGAATTAGAAGACCATGAAGAACGCAGC	214
Qy 1897	aggaaggagaatttagtattaagggagaagagaaaccaacgcggaagagagcgaagaag	19556
Db 215	ATGATGATGATTTTGAAATTAGAACAACGCCAGGAAGAGATGATGATCAAGAGAAGAAAGATG	274
Qy 1957	gagagccaacagaaggagaagtgcacaagaagaatttagaggccaactccagagacgatt	2016
Db 275	AAGAAGAAGAAGATGAAGNAGAAGATGAAGAAGATGAAGAAGACCATGATGAAGAAG	334
Qy 2017	tcgaattagaagaaccacagagagaagaatagaagaacacgcgtagaaggcgagaagaactg	2076
Db 335	ACGATGATGAAGAAGACCGATGATGAAGAAGACGATGATGAAGAAGACCGATGATGAAGAAG	394
Qy 2077	cagaaggagaagaagtggagaaggtacctcagaagtagaagaagtggagaaggtacctctg	2136
Db 395	ACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAGACGATGATGAAGAAG	454
Qy 2137	cagaagttagaagaagtggagaaggttaccaagaagtagaagaaggtaccgcgcagaagtag	2196
Db 455	ACGATGATGAAGACGACGATGATGATGAAGACGATGATGAAGAAGATGATGAAGATG	514
Qy 2197	aagaagtggagaaggtaccagaagaagtggagaaggtaccagaagaagtggagaaggtac	2256
Db 515	ATGATGATGAAGATGATGAATTTCAATTAGAAGACGATGATGAAGATGATGAAGATG	574
Qy 2257	cagaagaagtggagaaggtaccagaagaagtggagaagtggaagaagtagaagaagtag	2316
Db 575	ATGATGATGAAGATGACGATGATGATGAAGAAGATGATGATGAAGAAGATGATGATGAAG	634
Qy 2317	aggTaccagcggtagtagaagtagaagttaccgcggtagtagaagaagaggtgccagaag	2376
Db 635	AAGATGATGATGAAGAAGATGATGATGAAGATGACGATGATGATGAAGATGACGATGATG	694
Qy 2377	aagttagaagaagaagaagaggaagaacacgtagagggaagaagtagttattcaattag	2436
Db 695	ATGAAGAAGATGATGATCACGATGATGAAGAAGACGATGAAGAAGACGATGGAGAAGCCATC	754
Qy 2437	taataccatoggagaagaatatatacattagacaaccaaagaagacyaat	2488
Db 755	GAGATGACGATCGAGAAGACCATTTGACAAGACCGATGAAGATGAAGATGACGATGAAGT	806

RESULT	26	
AZ683582		
LOCUS	891 bp	DNA
DEFINITION	ENTKK47TR Entamoeba histolytica sheared DNA	linear GSS 14-DEC-2000
	genomic, DNA sequence.	Entamoeba histolytica
ACCESSION	AZ683582	
VERSION	AZ683582.1	GI:11820728
KEYWORDS	GSS.	
SOURCE	Entamoeba histolytica.	
ORGANISM	Entamoeba histolytica	
	Eukaryota; Entamoebidae; Entamoeba.	
REFERENCE	1 (bases 1 to 891)	
AUTHORS	Loftus, B., Van Aken, S. and Fraser, C.	
TITLE	Determination of clone end sequences from Entamoeba histolytica	
	HM1:IMSS sheared DNA library	
JOURNAL	Unpublished (2000)	
COMMENT.	Contact: Brendan J Loftus	
	Department of Eukaryotic Genomics	
	The Institute for Genomic Research	

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlouftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library

Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 16
High quality sequence stop: 694.

Location/Qualifiers
1. .891

FEATURES

source

/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

400 a 55 c 291 g 145 t

BASE COUNT
ORIGIN

Query Match 5.3%; Score 178; DB 12; Length 891;
Best Local Similarity 54.2%; Pred. No. 6.6e-14;
Matches 361; Conservative 0; Mismatches 305; Indels 0; Gaps 0;

QY 1790 tgtacaagttaccagtgccagtagggcccgcaagagagtgccacggaagaattgatgca 1849

Db 58 TGAAGATGATGAATTTGCAATTAGAACGCGATGATGAAGATGATGATGAAGATGATGA 117

QY 1850 actccaagaggcagattcgaattagaagaactcagaagctccagaggaagagaatt 1909

Db 118 TGAAGATGACGATGATGATGAAGAACGATGATGATGAAGAACGATGATGATGAAGAACGATGA 177

QY 1910 agtattagaaggaggaaggaacccaggaagagcgaagagcgaagagagagcgaagga 1969

Db 178 TGATGAAGAGATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGA 237

QY 1970 aggagaagtgccaggaagaagaattagaggcaactccaggagcagatttcgaattagaaga 2029

Db 238 AGATGATGATGATGATGATGAAGAACGCGATGAAGAACGCGATGAAGAACGCGATGAAGAACG 297

QY 2030 accaagaggaagaagtagaagaacccgtagaggcggaagaacactgcagaagagaaga 2089

Db 298 CGATGAAGAACGCGATGAAGAACGCGATGAAGAACGCGATGAAGAACGCGATGAAGAACG 357

QY 2090 agtgggaaggtaccctgcagaagtagaagaagtgggaagaggtaccctgcagaagtagaaga 2149

Db 358 AGACGATGATGATGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGA 417

QY 2150 agtgggaaggtaccaggaagaagtagaagaaggtaccgcgagaagtagaagaagtggaaga 2209

Db 418 TGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGA 477

QY 2210 ggtaccagagaagaagtggaagaggtaccagagaagtggaagaggtaccagagaagaagtga 2269

Db 478 TGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGA 537

QY 2270 agaggtaccagagaagaagtggaagaagtggaagaagtagaagaagtagaggtaccagcgg 2329

Db 538 TGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGATGAAGAACGCGATGA 597

QY 2330 agtagaagtagaagttaccagcgtagtagaagaagagtgccagagaagtagaagaaga 2389

Db 598 TGATGAAGAACGATGAAGATGATGAAGATGATGATGAAGAACGATGAAGAACGATGA 657

QY 2390 agaagaaggggaagaccagtagtagaagaagaagatgtattacaattatgattaccatcgga 2449

Db 658 TGAAGACGACGATGAAGACGACGATGAAGACGACGATGAAGACGACGATGAAGACGACGA 717

QY 2450 agaaga 2455

Db 718 TGAAGA 723

RESULT 27

LOCUS BH148582 942 bp DNA linear GSS 27-AUG-2001
DEFINITION ENTREP62TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION BH148582
VERSION BH148582.1 GI:15308029
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 942)
AUTHORS Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica HM1:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjlouftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 15
High quality sequence stop: 794.
Location/Qualifiers
1. .942

FEATURES

source

/organism="Entamoeba histolytica"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Borell, Oxford University Press, 1999)."

BASE COUNT 399 a 84 c 240 g 219 t
ORIGIN

Query Match 5.3%; Score 176; DB 12; Length 942;
Best Local Similarity 55.3%; Pred. No. 1.2e-13;
Matches 363; Conservative 0; Mismatches 290; Indels 3; Gaps 1;

QY 1855 aaagagcagatttcgattagaaggaactcagaagctccagaggaaggaattatgat 1914

Db 80 ATGATGAAGATGATGATTCCAATTAGACGAAGAAGATGATGAAGAAGATGAAGAG 139

```
Qy 1915 tagaagaggaaggaacccaacggaagaagccgaagcgaaggaaggaagcgaacagaagag 1974
Dy 140 ATGAAGAAGATGAAGAAGATCAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 199
Qy 1975 aagtgcagaagaagaattagagcgaactccagaggaagcatttgaattagaagaaccaa 2034
Dy 200 AAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATG 259
Qy 2035 cagagaagaagtgaagaacccctagaggcgaagaagcgaagaactgcagaaggagaagaattg 2094
Dy 260 AAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG 319
Qy 2095 aagaggctacctgcagaagtagaagaagtgggaagggtacctgcagaagtagaagaagtgg 2154
Dy 320 ACGAAGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 379
Qy 2155 aagagggtaccagaagaagtagaaggtaccgcgagaagtagaagaagtagaagaaggtac 2214
Dy 380 AAGACGAGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGA 439
Qy 2215 cagaagaagtgaagaggtaccagaagaagtgggaagggtaccagaagaagtagaagaag 2274
Dy 440 AATATGATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 499
Qy 2275 taccagaagaagtgggaagaagtgggaagaagtagaagaagtagaaggtaccagcgtagt 2334
Dy 500 AATTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 559
Qy 2335 aagtagaagtagcagcgtag---tagaagaagaaggtgcagaagaagtagaagaagaag 2391
Dy 560 ATCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 619
Qy 2392 aagaagaggaagaccagtagaagaagaagtgtattacaattagtaataccatcggaag 2451
Dy 620 AAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 679
Qy 2452 aagatatacaattagacaacccaagaagaagcgaattaggctcgtggaatttatct 2507
Dy 680 CAAAGTCAGAGGTTTCATCCATTAGATAAAATTGAATCAGCAGACGACTTCTTGTCT 735
```

```
RESULT 28
AZ550718/c
LOCUS
DEFINITION ENTEM36TF Entamoeba histolytica Sheared DNA linear GSS 14-NOV-2000
genomic, DNA sequence.
ACCESSION AZ550718
VERSION AZ550718.1 GI:11176019
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 879)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
The Institute of Eukaryotic Genomics
Department for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 25
High quality sequence stop: 801.
Location/Qualifiers
1..879
/organism="Entamoeba histolytica"
```

```
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/notes="Vector: pPOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT 193 a 219 c 72 g 395 t
ORIGIN

Query Match 5.3%; Score 175.8; DB 12; Length 879;
Best Local Similarity 55.5%; Pred. No. 1.3e-13;
Matches 339; Conservative 0; Mismatches 272; Indels 0; Gaps 0;

Qy 1822 aagaagtcgaacggaagaattgatgcactccaagaggagcatttcgaattagaagaa 1881
Dy 681 ACGTAGAGCTTTTGAAGAAGAAATAATGATGAAGATGATGAATTCCAATTAGACGAAG 622
Qy 1882 ctgcagaagctccagagaaggaattagattagtaggaaggaaggaaggaaggaag 1941
Dy 621 AAGAAGATGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 562
Qy 1942 aagagccaagagagggagccacagaagaagtgccagaagaagaattagaggcaa 2001
Dy 561 AAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATG 502
Qy 2002 ctccagaggcagatttcgaattagaagacccaagaggaagaagtagaagaacccgtag 2061
Dy 501 AAGAAGATGAAGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAG 442
Qy 2062 agggcagaagaactgcagaaggaagaagtggaaagggtacctgcagaagtagaagaag 2121
Dy 441 ATGAAGAAGATGAAGAAGATGAAGAAGACGAAGAAGACGAAGAAGATGAAGAAGATGAAG 382
Qy 2122 tggaagagggtacctgcagaagtagaagaagtggaaagggtaccagaagaagtagaag 2181
Dy 381 AAGATGAAGAAGATGAAGAAGATGAAGAAGACGAAGAAGACGAAGAAGACGAAGAAGACG 322
Qy 2182 taccgcagaagtagaagaaggtgggaaggtaccagaagaagtggaagaggtaccagaag 2241
Dy 321 AAGAAGACGAAGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAG 262
Qy 2242 aagtgggaaggtaccagaagaagtggaaggtaccagaagaagtggaagaggtggaag 2301
Dy 261 ATGCTGTTGCCGAGCTCTTTGCTGCTGAATTTGATGATGATGATGATGATGATGATGATG 202
Qy 2302 aagtagaagaagtagaggtaccagcggttagtagaagtagaagtagaccagcggttagtag 2361
Dy 201 AAGCTCTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 142
Qy 2362 aagaggtgccagaagaagtagaagaagaagaaggaaggaaggaaggaaggaaggaag 2421
Dy 141 CAGACTTTTTTGAAGACGAAGAAGATGAAGAAGATGAAGAAGATGAAGAAGATGAAGA 82
Qy 2422 atgtattacaa 2432
Dy 81 AATTAACCTCAA 71

RESULT 29
BM170146
LOCUS
DEFINITION EST572669 PyBS Plasmodium yoelii yoelii cDNA clone PYC0B91 5' end,


```

mRNA sequence.
BM170146
BI170146.1 GI:17303378
EST.
Plasmodium yoelii yoelii.
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
1 (bases 1 to 832)
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADF.
Location/Qualifiers
1..832
/organism="Plasmodium yoelii yoelii"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCQB91"
/clone_lib="PyBS"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidinium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridAP arms directionally using EcoRI-XhoI cleaved arms
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridAP vector and plasmid DNA
isolated."
BASE COUNT 330 a 154 c 207 g 141 t
ORIGIN

Query Match 5.3%; Score 175.2; DB 10; Length 832;
Best Local Similarity 58.2%; Pred. No. 1.5e-13;
Matches 328; Conservative 0; Mismatches 233; Indels 3; Gaps 1

QY 1969 gaattagaaggaactgcagaagctccagagagaaggaattagttatagaaggaagga 1928
|||||
Db 267 GAAGCACCAGTAGAAGCTGTAAGCACCAGTGAAGCTGAAGCACCAGTGAAGTAGAAGGA 326
|||||
QY 1929 gaaccaacgaagagcagaagagaagcacaacagaaggaaggaagtgcagaagaa 1988
|||||
Db 327 CCCGTAGAAGTAGAAGGAGCCCGTGAAGCTGAAGTACCAGTAGAAGCTGAAGTACCAGTA 386
|||||
QY 1989 gaattagaggcaactccagagacgatttcgaattatagaagcaaccacaggagaagta 2048
|||||
Db 387 GAAGCTGAATCACCAGTAGAAGATGAAGTACCAGTAGAATCAGAGTAGTAGTTGTAGAACT 446
|||||
QY 2049 gaagaacccgtagaggcgagaactgcagaaggaagaagtggaaggaaggtacctgca 2108
|||||
Db 447 GAAGCACCCTGTAAGAACTGAGGTACCAGCAAGTACCAGCAAGTACGCTGTGTAAGCTAGAAGT 506
|||||

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 10:11:49 ; Search time 72.28 Seconds
(without alignments)
11340.329 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagtaacaa.....tcttataaaataataattc 3337

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2.6/prodata/1/lna/5A_COMB.seq:*
2: /cgn2.6/prodata/1/lna/5B_COMB.seq:*
3: /cgn2.6/prodata/1/lna/6A_COMB.seq:*
4: /cgn2.6/prodata/1/lna/6B_COMB.seq:*
5: /cgn2.6/prodata/1/lna/PCUS_COMB.seq:*
6: /cgn2.6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	1 US-08-072-610-1	Sequence 1, Appli
2	3337	100.0	3337	2 US-08-719-822B-1	Sequence 1, Appli
3	3337	100.0	3337	4 US-09-092-458-1	Sequence 1, Appli
4	199	6.0	3489	2 US-08-728-323A-1	Sequence 1, Appli
5	199	6.0	32207	2 US-08-770-379-20	Sequence 20, Appli
6	199	6.0	32207	4 US-08-757-669A-20	Sequence 20, Appli
7	199	6.0	32207	4 US-09-230-371A-20	Sequence 20, Appli
8	175	5.2	3095	6 5231168-1	Patent No. 5231168
9	170.2	5.1	5361	4 US-08-973-462-2	Sequence 2, Appli
10	170.2	5.1	6152	4 US-08-973-462-1	Sequence 1, Appli
11	167.2	5.0	2580	3 US-09-050-863-2	Sequence 2, Appli
12	167.2	5.0	2580	4 US-09-359-081-2	Sequence 2, Appli
13	167.2	5.0	5452	2 US-09-130-114-1	Sequence 1, Appli
14	167.2	5.0	9800	4 US-08-910-647-1	Sequence 1, Appli
15	167.2	5.0	10596	1 US-07-884-811-15	Sequence 15, Appli
16	167.2	5.0	10596	1 US-07-885-971-15	Sequence 15, Appli
17	167.2	5.0	10596	1 US-08-087-783A-15	Sequence 15, Appli
18	167.2	5.0	10596	1 US-08-194-088B-15	Sequence 15, Appli
19	167.2	5.0	10596	1 US-08-194-087-15	Sequence 15, Appli
20	167.2	5.0	10596	5 PCR-US93-04648-15	Sequence 15, Appli
21	147.2	4.4	9551	1 US-08-056-200-93	Sequence 93, Appli
22	147.2	4.4	9551	2 US-08-800-644-93	Sequence 93, Appli
23	145	4.3	696	4 US-09-461-697-193	Sequence 193, App
24	145	4.3	699	4 US-09-461-697-191	Sequence 191, App
25	145	4.3	717	4 US-09-461-697-189	Sequence 189, App
26	145	4.3	774	4 US-09-461-697-187	Sequence 187, App
27	145	4.3	819	4 US-09-461-697-185	Sequence 185, App

28	145	4.3	1669	4 US-09-461-697-184	Sequence 184, App
29	141.8	4.2	7218	1 US-08-232-463-14	Sequence 14, Appli
30	138	4.1	9636	1 US-08-323-170B-1	Sequence 1, Appli
31	138	4.1	9636	4 US-08-954-441-1	Sequence 1, Appli
32	133.4	4.0	16442	3 US-08-781-891-208	Sequence 208, App
33	127.2	3.8	543	6 5273901-6	Patent No. 5273901
34	124	3.7	1276	4 US-09-177-325-2	Sequence 2, Appli
35	124	3.7	1276	4 US-09-411-812A-2	Sequence 2, Appli
36	124	3.7	1276	4 US-09-590-113-2	Sequence 2, Appli
37	124	3.7	3279	4 US-08-446-137B-1	Sequence 1, Appli
38	117.4	3.5	1891	4 US-08-973-462-3	Sequence 3, Appli
39	116.6	3.5	533	6 5482709-5	Patent No. 5482709
40	115.6	3.5	3211	2 US-08-574-959A-8	Sequence 8, Appli
41	115.6	3.5	3211	4 US-09-357-014-8	Sequence 8, Appli
42	115.6	3.5	3901	2 US-08-574-959A-6	Sequence 6, Appli
43	115.6	3.5	3901	4 US-09-357-014-6	Sequence 6, Appli
44	110.8	3.3	5433	3 US-08-929-329-1	Sequence 1, Appli
45	100.6	3.0	2129	4 US-08-975-762-39	Sequence 39, Appli

ALIGNMENTS

RESULT 1
US-08-072-610-1
; Sequence 1, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,610
; FILING DATE: 19930602
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/07686
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: pWMB3.3.1
US-08-072-610-1

Query Match 100.0%; Score 3337; DB 1; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;				
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	gaattccggttaagtaaacactatggttcgtatctatataataaccttaattttatc	60	
Db	1	GAATTCGGTAAAGTAACAACACTATGGTTTCGTATCTATATATAACCTTACTAATTTATC	60	
Qy	61	ttttgctttcttttaattcatcttcaacagtaagataaaaaataatctataaaaactgc	120	
Db	61	TTTTTGCTTTTCTTTAATTCATGCTTCAACAGTAAGATAAATAATATATATAAACTGC	120	
Qy	121	tatatatacatatattcatatagtggtcatttgtgaattgcgatcatcttaaaatttcgta	180	
Db	121	TATATATACATATATATCAATAAGTGGCATTTGTGAATTCGGATCATTTAAATTTACGTA	180	
Qy	181	aaacaataattgaaaaaatt	240	
Db	181	AAACAATAATTGAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTGTCTACAGACGATTTAG	240	
Qy	241	aattggaaaatgctctctgatgattgttagaggtgagggatccttcaaacgacgggttag	300	
Db	241	AATTGGAAAATGCTTCTGATGATGTTGTAGAGGTGGAGGATCCTTCAACGACGGTTTAG	300	
Qy	301	aattagaagaggaaaatttttgatgagaattcaggtgatgatgaactcttttagtgcta	360	
Db	301	AATTAGAAGAGGAAAATTTTGTATGAGAAATTCAGGTGATGATGAACCTCTTTTAGATGCTA	360	
Qy	361	ccccgaagatgactttgccttaacagatttgccaatggaagcagatgaggaagttcaacg	420	
Db	361	CCCCGAAGATGACTTTGCCCTTAACAGATTTGCCAATTTGAAGCAGTGAAGAGTCAACG	420	
Qy	421	aaacgttagatgaggtgaaatcattaggagaggtttccactgaagatatgaaacagaag	480	
Db	421	AAACGTTTAGATGGAGGTGAATCAATTAGGAGAGGTTTCCACTGAAGATATGGAACAGAA	480	
Qy	481	atggtcaacagatgatacagaacagaagaagactactcctggtgatataagaaggagaag	540	
Db	481	ATGGCTCAACAGATGATACCGGAACAGAAAGAGGACTTACCCTGGTGATATGGAAGGAGA	540	
Qy	541	aagaagctggcgtatagaagcagggggaagaagctggtgatttggagcagggggaagaaa	600	
Db	541	AAGAAGCTGGCATATGGAAGCAGGGGAAGAAGCTGGTGATTTGGAAGCAGGGGGAAGAA	600	
Qy	601	ctggcgtattggaagcagggggaactggcgtatttggagcaggggaagagcgtggtg	660	
Db	601	CTGGCGATTTTGGAGCAGGGGAAGAACTGGCGATTTTGGAGCAGGGGGAAGAGCTGGTG	660	
Qy	661	atttggagcaggggaagaactggcgatttgaagcagggggaagaactggagatgcgg	720	
Db	661	ATTTGGAGCAGGGGAAGAACTGGCGATTTTGGAGCAGGGGGAAGAACTGGAGATGCGG	720	
Qy	721	aaactgaagagagcaactggagatgcggaactggaactgaaatggagcaactgtgtatgtg	780	
Db	721	AAACTGAAGAAGGAGCAACTGGAGATGCGGAACCTGAAATTTGGAAGCACTGTGTATGTAG	780	
Qy	781	acacagaagatagttcagctgaaggagcagaaaaagtacatgttctctgctcaagaaaaatg	840	
Db	781	ACACAGAAGATAGTTCAGCTGATGGAGCAGAAAAAGTACATGTCTCTGCTCAAGAAAAATG	840	
Qy	841	tacaactccgatagtaattgacctcttggaaagtatttggataaagatatattt	900	
Db	841	TACAACTCCGATAGTAATTGATGCCCTCTTTGGAAGTATTTTGGATTAAGATATATAATT	900	
Qy	901	ttgatcataaagatttcagccactattcgaacaaaatttggcgggttactgctaaac	960	
Db	901	TTGATCATATTAAGATTTCAGCCACTATTGCAACAAATTTGTGGCGGTACTGCTTAAC	960	
Qy	961	atgttcagggaagaatttcgaatgaacactgtaccattaccagtggcagaagacccg	1020	
Db	961	ATGTTACGGGACAGAAATTCGAATGAACCTGTACCATTACCAGTGGCAGAAAGACCCG	1020	
Qy	1021	cgaagtcacagcgggaagattagatgcctccagggatgacttcgattagatttta	1080	
Db	1021	TACCTGCAGNAAGTAGAAGAGTGAAGAGTACTCTCAGAAAGTAGAAGAACTGGAGAGG	1080	

Db	1021	CGCAAGTACCAGCGGAAGAAATTAGATSCCACTCCAGAGGATGACTTCGCATTAGATGTTA	1080	
Qy	1081	cagaactctccaggaagtagaattagtagtaagaagcgaactgaagaagaatcaa	1140	
Db	1081	CAGAACTCTCCGAGGAAGTAGAAATTAGTATGAAGGCACTGAAGAGAATCAA	1140	
Qy	1141	cggaaatgggaccaacggaaggaaggaacacccgaagaattagatgcactccagagagtg	1200	
Db	1141	CGGAAGTGGACCAACGGAAGAGGACCAACCGAAGAATTAGATGCCACTCCAGAGGATG	1200	
Qy	1201	gatttcgacttagcagaactgcagaagagagaacagaagaacgttagagggagaagaaa	1260	
Db	1201	GATTTCCGCAATTAGACAAACTGCAGAGGAGAAACAGAAACGTAGAGGGAGAGAAA	1260	
Qy	1261	cagaagaagctgcagaaggaagtagatcagaagaactccagaaggaagaagaagtag	1320	
Db	1261	CAGAAGAAGCTGCAGAAGGAGAGTATCAGAAGAACTCCAGAAGGAGAGAGAGTTAG	1320	
Qy	1321	aggcaactccagaggtatttcgcattagatggaactacattagaagaacccaagaaa	1380	
Db	1321	AGGCAACTCCAGAGGATGATTTGCGATTAGATGGAACCTACATTAGAGAAACCGAAGAA	1380	
Qy	1381	ctgcagaagagaagaacccgtagaggaagaacacccgtagaggaagaagaacccgtag	1440	
Db	1381	CTGCAGAAGGAGAAACCCGTAGAGGGAGAGAAACCCGTAGAGGGAGAGAAACCCGTAG	1440	
Qy	1441	agggaagaagctgcagaaggaagaagtagtaggccaactccagagagtagacttcc	1500	
Db	1441	AGSGAGAAGAGCTGCAGAAGGAGAGAGAGTTAGAGGCAACTCCAGAGGATGACTTCC	1500	
Qy	1501	aattagaagaaccatcagagaaggaaggggaaggaagaaggaaggaaggaaggaag	1560	
Db	1501	AATTAGAAGAACCATCAGGAGAAAGGAGGGAAGGAGAGGGAAGGGAAGGGAAGGAAG	1560	
Qy	1561	gagaacggttagtagcagtcagtagtgccgaacccgtagaagtagtagtactcctgctc	1620	
Db	1561	GAGAACGTTTAGTAGCAGTGCCAGTAGTGCCCACTAGTGCCCAACCCGTAGAGTAGTAGTACTCTGCTC	1620	
Qy	1621	agcctgtcaaaccaatgctcctccaacgcagatgaactttattcgttgatatttag	1680	
Db	1621	AGCCTGTCAAAACCAATGGTCGCTCCAACGCGCAGATGAAACTTTATTCGTTGATATCTTAG	1680	
Qy	1681	ataacgatttaacgtatgcagacattacatcctttgagccattatttaacaatacctca	1740	
Db	1681	ATAACGATTTAACGTATGCAGACATTACATCTCTTTGAGCCATTATTATAACAATCCTCA	1740	
Qy	1741	aggatcctgatgcagagagcgtgtacacgtaccatcaaggaagcaccctgttacaagtac	1800	
Db	1741	AGGATCCTGATGCAGGAGGCTGTAACTACCTACCAAGGAAGCAACCTGTACAAAGTAC	1800	
Qy	1801	cagtgagtagggccgcgcaagaagtgcacaacggaagaattgtgcactccaagagg	1860	
Db	1801	CAGTGGCAGTAGGGCCCGCCCAAGAGTGCCAACGGAAGAAATTTGATGCAACTCCAAGAGG	1860	
Qy	1861	acgatttcgaatttagaaggaactgcagaagctccagaggaaggaagatttagttagaag	1920	
Db	1861	ACGATTTCGAATTTAGAAGGAACCTGCAGAGCTCCAGAGGAAGGAGAAATTAGTATTAGAAG	1920	
Qy	1921	gagaaggaagaacacgcggaagaagcgaagaagaagagagccaaacagaagggagaagtgc	1980	
Db	1921	GAGAAGGAGNACCACCGAAGAAAGAGCCCAAGAGAGAGAGCCCAACAGAAGGAGAGTGC	1980	
Qy	1981	cagaagaagaatttagaggccaactccagagagcatttcgaattagaagaaccaacaggag	2040	
Db	1981	CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTCCGAATTTAGAAGAACCACAGGAG	2040	
Qy	2041	aagaagtagaagaaaacgtagagggcgaagaaactcagaaggaaggaagaagtggagaagg	2100	
Db	2041	AAGAAGTAGAAGAAAACCGTAGAGGGCGAAGAACTGCAGAAGGAGAGAAAGTGGAAAGAGG	2100	
Qy	2101	tacctcagaagtagaagaagtggagaggtacctcagaagtagaagaagtggagaagg	2160	
Db	2101	TACCTGCAGNAAGTAGAAGAGTGAAGAGTACTCTCAGAAAGTAGAAGAACTGGAGAGG	2160	

QY 2161 taccagaagaagtagaagaggtaccgcgagaagtagaagaagtggaagaggtaccagaag 2220
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCAGAGAGTACCAGAGTACCAGAGTACCAGAG 2220
QY 2221 aagtggaagaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccag 2280
Db 2221 AAGTGAAGAGGTACCAGAAGAAGTGAAGAGGTACCAGAAGAAGTGAAGAGGTACCAG 2280
QY 2281 aagaagtggaagaagtggaagaagtagaagaagtagaaggtaccagcggtagtagaagtag 2340
Db 2281 AAGAAGTGAAGAAGTGAAGAAGTGAAGAAGTGAAGAAGTGAAGAAGTGAAGAAGTGA 2340
QY 2341 aagtcacagcggtagtagaagaagaggtgccagaagaagtagaagaagaagaagaag 2400
Db 2341 AAGTACCAGCGGTAGTAGAAGAGAGGTGCCAGAAGAAGTAGAAGAAGAAGAAGAGG 2400
QY 2401 aagaaccagtagaggaagaagtagattacaattagtaataaccatcggaagaagatatac 2460
Db 2401 AAGAACCAGTAGAGGAAGAAGATGTATTACAATTAGTAATACCATCGGAAGAAGATATAC 2460
QY 2461 aattagacaaacaaagaagaacgaattaggtcctggaatttttatctatcatcgacatgc 2520
Db 2461 AATTAGACAACCAAGAAAGAGCAATTAGGCTCTGGAATTTTATCTATCATCGACATGC 2520
QY 2521 actaccaagacgttccaaaggaatttatggaagaagaagaactgcagtgatccat 2580
Db 2521 ACTTACCAGACGTTCCAAAGGANTTTATGGAAGAAGAAGAAACTGCAGTGATCCAT 2580
QY 2581 tgaaccagaagattttgcaaggaagattcaccaattcagaaatggctcacatctc 2640
Db 2581 TGAACCCAGAAGATTTTGAAGGAAGATTCACAATCTCAGAAATGGCTCACATTCATTC 2640
QY 2641 aagccctaaagcagactgggaacgattagaagtgagcttaataaggttagaagaagat 2700
Db 2641 AAGCCCTAGAAGCGACTGGGAACGATTAGAAGTGAGCTTAAATAGGCTAGAGAAGAT 2700
QY 2701 ggaatgaaacaaagaataaagaatggctggctggctggctggctggctggctggctggct 2760
Db 2701 GGATGGAACAAAGAAATAAAGATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 2760
QY 2761 cagaatatagtcataattcaacaaaggaaagaccagctggcttggaagaacagagat 2820
Db 2761 CAGAAATATAGTCAAAATTTCAACAAAAAGAAAGGACCCAGCTGGTTTGCAGAAAAAGAG 2820
QY 2821 ggaagcagcagaagtagaagaatggtttaaagcagaagtcacaaatcccaattgattcac 2880
Db 2821 GGAGCCAGCAGAANTGGAAAAATGGTTTAAACGAGAAGTCAATCCCAATTTGATTTCAC 2880
QY 2881 acttgaaaaatggatgaacgacactcattcccaatttttaaaattcttgtaagaagata 2940
Db 2881 ACTTGAAAAATGGATGAACGACACTCATTCCCAATTTTAAATTTCTTGTAAGAAGATA 2940
QY 2941 tgcacaatttgaaaaacaagaacaaagaaatggtaaatcactcggaagaagaag 3000
Db 2941 TGTCACAATTTGAAAAACAAGAAACAAAGAAATGGTTAATGAATCACTCGAAAAAGAACG 3000
QY 3001 aacggggttatggtctgaatcatttgaaattatgaccatcaaaattattaaatgtg 3060
Db 3001 AACGGGGTTATGGTCTGAAATCATTGAAAGTTATGACCACATCAAAATTTTAAATTTGG 3060
QY 3061 ctaagagtcgagaatggtaccgtgccaatctctaataaataagaaaaagaagaactca 3120
Db 3061 CTAAAGAGTCAGAATGGTACCGTGCCAATCTTAATAATAGAGAAAGAGAACTCA 3120
QY 3121 tgaataatgggtttcccttaaaagaacgaataatttagagcaaaagaatggaaaaaatggact 3180
Db 3121 TGAATATGGGTTTCCCTTAAAGAAAGCAATTTTATAGSACAAAGAAATGGAAAAATGGACT 3180
QY 3181 catggaaaaaagttaaaatttttgggttcaattcaatggtgacacattttctgaaaa 3240
Db 3181 CATTGAAAAAAGTTAAATTTTGGTTCATTCATGTTGACACATTTTCTTCTGAAAAA 3240

QY 3241 cgctaaccaagaagaatggaatcaattgttaataagaaataaaagtttgaattatagaa 3300
Db 3241 CGCCTAACCAAGCAAGATGGAATCAATTTGTTAATCAATAAAGCTTTCATATATAGAA 3300
QY 3301 aaagaacagagattctctctataaaaaataaataattc 3337
Db 3301 AAAAGAAGACAGATTATTCTCTTATATAAAATAAATAATTC 3337
RESULT 2
US-08-719-822B-1
; Sequence 1, Application US/08719822B
; Patent No. 5874527
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/719,822B
; FILING DATE: 09/30/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3337 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PVM3.3.1
; US-08-719-822B-1

Query Match 100.0%; Score 3337; DB 2; Length 3337;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 gaattccggtaaaagtaacaaactatggttcgtatctatataaaccttactaattttatc 60
Db 1 GAATTCGGTAAAGTAACAACTATGTTTCGTATCTATATATAACCTTACTAATTTATC 60
QY 61 ttttgctttcttttaatttcatttcacagtaagataaaaaataatctctataaaactgc 120
Db 61 TTTTGTGTTTCTTTAATTCATGCTTCAACAGTAAGATAAAAAATAATCTATAAAACCTGC 120
QY 121 tatataacatatatttcacaaagtggttcatttgcatttgcatttgcatttgcatttgcatt 180
Db 121 TATATACATATATATATTCATAAGTGGCATTTGTGAATTCGATCATTTTAAATTTACGTA 180

QY 361 cccccgaagatgacttgccttaacagattgccaattgaagcagatgaggaagtcaacg 420
Db 361 CCCCGAAGATGACTTTGCGCTTAACAGATTGCCAATTGAAGACGATGAGGAAGTCAACG 420
QY 421 aaacgttagatgaggtgaaatcattagagagaggtttccactgaagatatggaacagaag 480
Db 421 AAACGTTAGATGGAGGTGAATCATATTAGAGAGGTTTCCACTGAAGATATGGAACAGAAG 480
QY 481 atgctcaacagatgatacagaaacagaagaagactacacctggtgatatggaagagaag 540
Db 481 ATGCTCAACAGATGATACGGAAACACAGAGAGGACTACCTGGTGATATGGAAGCAGAAG 540
QY 541 aagaagctggcgatatgaaagcaggggaagaagctggtgatattggaagcaggggaagaaa 600
Db 541 AAGAAGCTGGCGATATGGAAGCAGGGGAAGAAGCTGCTGATTTGGAACGAGGGGAAGAAA 600
QY 601 ctggcgatttggaaagcaggggaagaactggcgatttggaaagcaggggaagaagctggtg 660
Db 601 CTGGCGATTTGGAAGCAGGGGAAGAAACTGGCGATTTGGAAGCAGGGGAAGAACTGGTG 660
QY 661 atttggaaagcaggggaagaactggcgatttggaaagcaggggaagaagctggagatcgg 720
Db 661 ATTTGGAAGCAGGGGAAGAACTGGCGATTTGGAAGCAGGGGAAGAACTGGAGATCGG 720
QY 721 aaactgaagaagggcaactggagatcggaactggaactgaaactggaactggtgtatgtag 780
Db 721 AAACCTGAAGAAGGAGCAACTGGAGATCGGGAACCTGAAATGGAGCAACTGTGTATCTAG 780
QY 781 acacagaagatagttcagctgtagtgaagcagagaagaagtacatgttccctgctcaagaagaatg 840
Db 781 ACACAGAAGATAGTTCAGCTGATGGAGCAGAAAAGTACATGTTCCTGCTCAAGAAAATG 840
QY 841 tacaacctgcgcatagtaactgacctcttggaaatttggataaaagataataatt 900
Db 841 TACAACCTGCGATAGTAAATGATGCCCTCTTTGGAAATTTTGGATAAAGATATAATTT 900
QY 901 ttgatcatattaaagatttcagcgccactattcgaaacaaatttggcgggtactgctaaac 960
Db 901 TTGATCATATTAAAGATTTCAGCGCCACTATTTCGAACAAATTTGTGGCGGGTACTGCTAAAC 960
QY 961 atgttcaggacaagaattcccaatgaacctgtaccattaccagtgccagagaagcccg 1020
Db 961 ATGTTACGGGACAGAAATTCGCAATGAACCTGTACCATTACCAAGTGCGCAGAGGCCG 1020
QY 1021 cgcaagtaccagcgggaagaattagatgccactccagaggtgacttcgcattagatgta 1080
Db 1021 CGCAAGTACCAGCGGAAGAAATTAGATGCCACTCCAGAGGATGACTTCGCATTAGATGTTA 1080
QY 1081 cagaactcccgaggaagtagaattagattagatgaagggcaactgaagaagaatcaa 1140
Db 1081 CAGAATCTCCGAGGAAGTAGAATTAGTATTAGATGAAGGCAACTGAAGAAGAATCAA 1140
QY 1141 cggaaagtggaccacgaagaagcccaacgaagaattagatgacctccagagagatg 1200
Db 1141 CGGAAGTGGGACCAACCGAGAGAGCCACCAGGAATTTAGATGCCACTCCAGAGGATG 1200
QY 1201 gatttcgattagacgaactgcagaagggagaacagaagaactgtagaggggaagaaa 1260
Db 1201 GATTTTCGATTAGACGAACCTGCAGAGGAGAAACAGAGAACCTGAGGGGAGAAGAAA 1260
QY 1261 cagaagaagctgcagaagggagaadtatcagaagaactccagaagaagtagaagagtag 1320
Db 1261 CAGAAGAAGCTGCAGAGGAGAGATATCAGAAGAATCTCCAGAAGGAGGAAGAGGTAG 1320
QY 1321 aggcaactccagagatgatttcgattagatggaactacattagaagaacccaagaaa 1380
Db 1321 AGGCAACTCCAGAGATGATTTCGATTAGATTGGAATACATTAGAGAAGAACCGAAGAAA 1380
QY 1381 ctgcagaaggaagaacccgtagaggagaagaacccgtagagggaagaagaacccgtag 1440
Db 1381 CTGCAGAAGGAGAAACCGTAGAGGGAGAGAAGAACCGTAGAGGGGAGAGAAGAACCGCTAG 1440

QY 1441 agggagaagaagctgcagaagggagaagaagtttagaggccaactccagagagatgacttcc 1500
Db 1441 AGGCAGAAGAAGCTGCAGAAGGAGAGAGTTAGAGGCAACTCCAGAGGATGACTTCC 1500
QY 1501 aattagaagaacatcaggaagaaggaaggaaggaaggaaggaaggaaggaaggaag 1560
Db 1501 AATTAGAAGAACCATCAGGAGAGGAGGGAAGGGAAGGAGGGAAGGGAAGGAGAAG 1560
QY 1561 gagaagcgttagtagcagtgccagtagtggccgaaccggtagaagtagtagtactctctc 1620
Db 1561 GAGAAAGCTTTAGTAGCAGTGCCAGTAGTGCCGAACCCGTAGAAGTAGTGACTCTCTC 1620
QY 1621 agcctgtcaaacccaatggtcgtccaaacggcagatgaacttttctgtgatatcttag 1680
Db 1621 AGCCTGTCAAAACCAATGCTGCTCCAACGGCAGATGAACCTTTATTCGTTGATATCTTAG 1680
QY 1681 ataacgatttaacgtatgcagacattacatccttttgaccattatttaacaacatctca 1740
Db 1681 ATAACGATTTAACTGATGCAGACATTACATCCTTTTGAGCCATTATTTAAACAATCTCA 1740
QY 1741 aggtacctgtagcaggaagcgtgtaacagtaccatcaaggaagcaccctgtacaagtac 1800
Db 1741 AGGATCCTGATGCAGGAGAGGCTGTACAGTACCATCAAGGAAGCACTGTACAAGTAC 1800
QY 1801 cagtggcagtagggccgcgcaagaagtgcacaacggaagaattgatgcaactccaagagg 1860
Db 1801 CAGTGGCAGTAGGGCCGCGCAGAACTGCAAGTGCACCGAAGAAATTTGATGCAACTCCAAGAGG 1860
QY 1861 acgatttcgaattagaagaactgcagaactccagaggaaggaagattagattagaag 1920
Db 1861 ACGATTCGAATTAGAAGGAACCTGCAGAGCTCCAGAGGAGGAGAATTAGTATTAGAAG 1920
QY 1921 gagaagagaaacaaacggaagaagcgaagagagagagagcccaagaggaaggaagtgc 1980
Db 1921 GAGAAAGGAGAAACCAACGSAAGAGAGCCAAAGAGAAGGAGGCCAACAGAGGAGAAGTGC 1980
QY 1981 cagaagaagaattagaggcaactccagagagcgaatttcgaattagaagaaccaacagagg 2040
Db 1981 CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTTCGAATTTAGAAGAACCAACAGAGG 2040
QY 2041 aagaagtagaagaacccgtagagggcgaagaactgcagaaggaagaagtggaagagg 2100
Db 2041 AAGNACTAGAAGAAACCGTAGAGGCCGAAGAACTGCAGAAGGAGAGAGAGTGGAAAGG 2100
QY 2101 taactgcagaagtagaagaagtggaaaggtacctgcagaagtagaagaagtggaagagg 2160
Db 2101 TACCTGCAGAAGTAGAAGAAGTGGAAAGGTACCTGCGAAGATAGAAGAAGTGGAAAGG 2160
QY 2161 taccagaagaagtagaagaggtaccgcagaagtagaagaagtggaaggtaccagaag 2220
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCCGCAAGATAGAAGAGTGGAAAGGTACCAAGAG 2220
QY 2221 aagtggaaaggttaccagaagaagtggaaagagggtaccagaagaagtggaagaggtaccag 2280
Db 2221 ANGTGGAAGAGGTACCAGAGAAGAGTGGCAAGGTTACCAGAAGAGTGGAAAGGTACCAG 2280
QY 2281 aagaagtggaagaagtggaagaagtagaagaagttagagggtaccagcgggttagtagaagtag 2340
Db 2281 AAGAAGTGGAAAGAGTGGAAAGATAGAGAAGGTAGAGGTACCAAGCGGTAGTAGAAGTAG 2340
QY 2341 aagtaccagcggtagtagaagaagaggtgccagaagaagtagaagaagaagaagaagagg 2400
Db 2341 AAGTACCAGCGGTAGTAGAGAAGAGGTGCCAGAAGAGTAGAAGAAGAGAAGAGAGG 2400
QY 2401 aagaacccagtagaggaagaagatgtattacaattagtaataaccatcggaagaagataac 2460
Db 2401 AAGNACCAGTAGAGGAAGAGATGTATTACAATTTAGTTAGTAAATACCATCGGAAGAGATATAC 2460
QY 2461 aattagacaaaccaaagaagaagaattaggctctggaatttttctatcatcatgcacatgc 2520
Db 2461 AATTAGACAAACCAAGAAAGACGAATTAGGCTCTGGAATTTTATCTATCATCTGACATGC 2520
QY 2521 actaccaagcagtttccaaaggaatttatggaagaagaagaagaactgcagtgatccat 2580

Db 2521 ACTACCAAGACGTTCCAAAGCAATTTATGGAAGAAGAAGAACTGCAGTGTATCCAT 2580
Qy 2581 tgaaccagaaattttgcaagaagatcacaatctacagaatgctcacattcattc 2640
Db 2581 TGAACCCAGAAGATTTGCAAGGAAGATTCACAAATCTACAGATGGCTCACATTCATC 2640
Qy 2641 aagccctagagggactgggaacgattagaagtgaagcttaataaagcctagagaaagat 2700
Db 2641 AAGCCCTAGAGGGGACTGGGAACGATTAGAAGTGAGCTTAATAAAGCTAGAGAAAGAT 2700
Qy 2701 ggatggaaacaaataaagaattggctggctggctggctggctggctggctggctggct 2760
Db 2701 GGATGGAACAAAGAAATFAAAGATGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGG 2760
Qy 2761 cagaatatagtcacaaatttcacaaaggaaggaagcagctggctggctggctggctggct 2820
Db 2761 CAGAAATAGTCAATTTCAACAAAGGAAGGACCCAGCTGGTTTGAGAAACGACAGT 2820
Qy 2821 ggagcagcagaaatggaaataatgggttaaaagcagaagtcacaaatcccaattgattcac 2880
Db 2821 GGAGCGACGAAATGGTAAAGAAATGGTTAAAGCAGAAGTCAAAATCCCAATTCATTCCAC 2880
Qy 2881 acttgaataatggaatgaacgacactcattcccaattttatttaaaattcttggtaagata 2940
Db 2881 ACTTGAAATAATGGATGAACGACACTCATTCCTCAATTTATTATAAATTTCTGTGAAAGATA 2940
Qy 2941 tgtcacaaatttgaacaaagaaacccaaagaatggttaaatgaatcactggaaagaacg 3000
Db 2941 TGTCACAATTTGAACAAACAGAAACCAAGAAATGTTAATGAATCACTGGGAAAGAACG 3000
Qy 3001 aacggggatttggttcgaatcatttgaagttagccacatcaaaattattaaattgg 3060
Db 3001 AACGGGTTATGTTCTGAATCATTTGAAGTTATGACCACATCAAAATTTATTAAATGTG 3060
Qy 3061 ctaagaatcagaatggtaccgtcccaatcctaataataatagaagaagaactca 3120
Db 3061 CTAAGATCGAGAATGTTACCGTGCCCAATCCTAATAATAATAGAGAAGAAAGAACTCA 3120
Qy 3121 tgaattggttctctaaagaaacgaatatttaggacaaagaatggaaataatggact 3180
Db 3121 TGAATGTTCTCTTAAGAAACCAAGAAATTTAGGCAAAAGAAATGGAAATGGACT 3180
Qy 3181 cattggaaaaagtaatttttggttcccaattcaatggtgtacacattttctgggaaa 3240
Db 3181 CATTGGAAAAAGTTAAATTTTGTGTTCAATTCATTCATTCATTCATTCATTCATTCATTC 3240
Qy 3241 cgcctaaccaggaagaatggaatcgaattggttaataaataaagaattggaattatagaa 3300
Db 3241 GCCTAACCAAGGAAGTAATGGAATTTGTTTAAATGAAATTAATGAAATTTGAAATTTATGAA 3300
Qy 3301 aaagaaacagattattctctataataataataattc 3337
Db 3301 AAAAGAACAGATTATTCTCTTATAAATAAATAATTC 3337

RESULT 4
US-08-728-323A-1
; Sequence 1, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas

CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/728,323A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 0575/52268/JPW/MSC/SKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-278-0400
TELEFAX: 212-391-0525
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3489 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3489
US-08-728-323A-1

Query Match 6.0%; Score 199; DB 2; Length 3489;
Best Local Similarity 46.9%; Pred. No. 2.5e-27;
Matches 685; Conservative 0; Mismatches 770; Indels 4; Gaps 2;
Qy 1011 gaagagccgcgcgaagtaccgcggaagaatttagatgccatccagagagtgacttcgca 1070
Db 1027 GAGGATCAGCAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 1086
Qy 1071 ttagatgtacagaatctccgcgagaagtagaatttagatttagatgaagagcaactgaa 1130
Db 1087 GAGGATCAGCAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 1145
Qy 1131 gaagaaatcaacgcgaagtgggaccacgaaggaagaccacccgaagaatttagatgccact 1190
Db 1146 GGAGGATCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1205
Qy 1191 ccagagatggatttcgatttagcagaactcgcagaaggaaggaaggaaggaaggaaggaag 1250
Db 1206 CGAGGAGGATCAGATGATGAGGACAAATGAGGACGAGGAGGATGACGAGGAGGAGGAGGAG 1265
Qy 1251 ggagaagaacacagaagaagctgcagaagagaagtagatcagaagaactccagaagagaa 1310
Db 1266 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1325
Qy 1311 gaagagtttagaggcgaactccagagagatgatttcgatttagatgggaactacattagaaga 1370
Db 1326 ACAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1385
Qy 1371 accgaagaactgcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1427
Db 1386 CTGCAGGAGGCCAACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1445
Qy 1428 gaagaaacccgttagagggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1487
Db 1446 CGAGGAGGCCAACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1505
Qy 1488 gaggatgacttcccaattagaagaaccatcagagaaggaaggaaggaaggaaggaaggaag 1547
Db 1506 GGAGGCCAACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1565


```
Db 20611 CCTCAGGAGCCACACAGCAGGAGGCCACAGCAGCAGGAGCCACAGCAGGAGGCCCT 20552
Qy 1428 gaagaaacccgtagaggagagaagctgcagaagagagaagaagtagtaggccaactcca 1487
Db 20551 GCAGGAGCCACACAGCAGGAGCCACAGCAGCAGGAGGCCCTGCAGGAGCCACACAGCA 20492
Qy 1488 gaggatgacttccaattagaaagaccacacaggaaggaagggggaaggaagagaa 1547
Db 20491 GGAGCCACAACAGCAGGAGCCACAGCAGCAGGAGGCCACAGCAGGAGCCACAGCAGCA 20432
Qy 1548 ggggaaggaaggaagagcgttagtagcagtagtccagtagtgccgaacccggtagaagta 1607
Db 20431 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGGCCACAGCAGGAGCCACAGCAGCA 20372
Qy 1608 gtgactcctgcagcctgtcaaaccaatgtgtcctcaacgagcagatgaactttattc 1667
Db 20371 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGGCCACAGCAGGAGCCACAGCAGCG 20312
Qy 1668 gttgatatttagataacgatttaacgtatgcagaaacattacatccctttgagccatttt 1727
Db 20311 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGGCCACAGCAGGAGCCACAGCAGCG 20252
Qy 1728 aacaaatcctcaaggtctctgtagcagagaggtctgaacagtagtaccatcaaggaagca 1787
Db 20251 GGAGCCACAGCAGGAGCCACAGCAGCAGGAGGCCACAGCAGGAGCCACAGCAGCA 20192
Qy 1788 cctgtacaagtaccagtgccagtagggccgcgcgaagaagtgcacaacggaagaattgatg 1847
Db 20191 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGAGGATGAGCAGCAGGATGAGCAGCAGCA 20132
Qy 1848 caactccaagagagcatttcgaattagaaggaactgcagaagctccagaggaagagaa 1907
Db 20131 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 20072
Qy 1908 ttagtattagaaggaaggaagacccaacgcgaagaagagccaagagagagagccaaca 1967
Db 20071 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 20012
Qy 1968 gaaggagaagtccagaagaagattagaggaactccagaggaacttcgaattagaa 2027
Db 20011 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19952
Qy 2028 gaaccaacaggaagagtagaagaacccgtagagggcgaagaactgcagaagagaa 2087
Db 19951 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19892
Qy 2088 gaagtggaaagggtacctgcagaagttagaagaagtggaaagggtacctgcagaagttagaa 2147
Db 19891 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19832
Qy 2148 gaagtggaaagggtaccagaagaagttagaagaaggtaccgcagaagttagaagaagtggaa 2207
Db 19831 GCAGCAGCAGCAGGATGAGCAGCAGCAGCAGGAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19772
Qy 2208 gaggtagcagaagaagtggaaagggtaccagaagaagtggaaagggtaccagaagaagtgg 2267
Db 19771 GCAGCAGGATGAGCAGCAGCAGCAGGAGGATGAGCAGCAGCAGGAGGAGGAGCAGCAGCA 19712
Qy 2268 gaagaggtaccagaagaagtggaaagaagttagaagaagttagaggtaccagcg 2327
Db 19711 GGAGGACAGCAGGAGGAGGATGAGGAGGAGCAGCAGGAGGATGAGGAGGATGAGGAGCAGCA 19652
Qy 2328 gtagtagaagttagaagtagcagcgggttagtagaagaagaggtgcagagaagaagtagaagaa 2387
Db 19651 GTTAGAGGAGCAGCAGCAGGAGGATGAGGAGGAGCAGCAGGAGGATGAGGAGGAGCAGCAGCA 19592
Qy 2388 gaagaagaagaaggaagaccagtagaagaagaagatgtattacaattagtaataccatcg 2447
Db 19591 GGAGTTAGAGGAGCAGCAGCAGGAGGATGAGGAGGAGCAGCAGGAGGATGAGGAGGAGCAGCA 19532
Qy 2448 gaagaagatatacaattag 2466
Db 19531 GCAGGAGTTAGAGGAGCAG 19513
```

RESULT 7

```
US-09-230-371A-20/c
; Sequence 20, Application US/09230371A
; Patent No. 6348586
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A
; APPLICANT: Russo, James J
; APPLICANT: Edelman, Isidore S
; APPLICANT: Moore, Patrick S
; TITLE OF INVENTION: UNIQUE ASSOCIATED KAPOSI'S SARCOMA VIRUS SEQUENCES AND
; FILE REFERENCE: 45185-G-PCT-US
; CURRENT APPLICATION NUMBER: US/09/230, 371A
; CURRENT FILING DATE: 1999-11-17
; PRIOR APPLICATION NUMBER: PCT/US97/13346
; PRIOR FILING DATE: 1997-07-22
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
; LENGTH: 32207
; TYPE: DNA
; ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-230-371A-20
```

```
Query Match 6.0%; Score 199; DB 4; Length 32207;
Best Local Similarity 46.9%; Pred. No. 4.2e-27;
Matches 685; Conservative 0; Mismatches 770; Indels 4; Gaps 2;
```

```
Qy 1011 gaagagccccgcaagtaccagcggaagaattagatgccactccagaggtgacttcgca 1070
Db 20970 GAGGATCAGCAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 20911
Qy 1071 ttgatgttacagaatctcccgaggaagtagaattagattagatgaagaggcaactgaa 1130
Db 20910 GAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 20852
Qy 1131 gaagaatcaacgcgaagtgggacccaacgcgaagaaggaaggaaggaaggaaggaag 1190
Db 20851 GGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGA 20792
Qy 1191 ccagaggtatggtatttcgcattagacgaactgcagaaggaaggaaggaaggaaggaag 1250
Db 20791 CGAGGAGGATGACGATGATGAGGACCAATGAGGACGAGGAGGATGACGAGGAGGAGGACAA 20732
Qy 1251 ggagaagaaacagaagaagctgcagaaggaaggaagtagatcagaagaactccagaagga 1310
Db 20731 GAAGGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20672
Qy 1311 gaagagtagaggcaactccagaggtatgttcgcattagatggaactacattagaagaa 1370
Db 20671 ACAGCAGCAGGAGCCACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20612
Qy 1371 accgaagaaactgcagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 1427
Db 20611 CCTGCAGGAGCCACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20552
Qy 1428 gaagaaacccgtagagggaaggaagctgcagaaggaaggaaggaaggaaggaaggaagga 1487
Db 20551 GCAGGAGCCACACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 20492
Qy 1488 gaggatgacttccaattagaagaaccatcaggaaggaaggaaggaaggaaggaaggaagga 1547
Db 20491 GGAGCCACAACAGCAGGAGCCACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20432
Qy 1548 ggggaaggaaggaaggaagcgttagtagcagtagtgcagtagtgccgaacccggttagaagta 1607
Db 20431 GGAGCCACAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 20372
Qy 1608 gtgactcctgcagcctgtcaaaccaatgtgtcctcaacgagcagatgaactttattc 1667
```

```

Db 20371 GGAGCCACAGCAGCAGGAGCCACAGAGCGGGAGCCACAGCAGCGGGAGCCCGCAGCAGCG 20312
QY 1668 gttgatctcttagataacgatttaacgtatcgagacattacatcccttttagccattattt 1727
Db 20311 GGAGCCACAGCAGCGGGAGCCACAGAGCGGGAGCCACAGCAGCGGGAGCCCGCAGCAGCG 20252
QY 1728 aaacaaatcctcaaggatcctgatcgagagagggctgttaacagtacatccatcaaaaggaagca 1787
Db 20251 GGAGCCACAGCAGCGGGAGCCACAGCAGCGGGAGCCACAGCAGCAGGATGAGCAGCAGCA 20192
QY 1788 cctgtacaagtaccagtgagtagggcccgcgcaagaagtgccccagcgaagaattgatg 1847
Db 20191 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 20132
QY 1848 caactccaagagacgatttcaatttagaagaactcgaaagctcgaaagctccagagaagagaa 1907
Db 20131 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 20072
QY 1908 ttagtattagaaggagaaggaaccaaagggaagagccaaagagaagagagcaaca 1967
Db 20071 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGCAGGATGAGCAGCAGCA 20012
QY 1968 gaaggagaagtcocagaagaagaatttagaggcaactccagagagacgatttccgaattagaa 2027
Db 20011 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19952
QY 2028 gaaccaacaggaagaagtagaagaacccgtagagggcgaaagaaactcgaaagagaa 2087
Db 19951 GGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19892
QY 2088 gaagtggagaggttaacctgcagaagttagaagaagtggagaggtacctgcagaagttagaa 2147
Db 19891 GCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19832
QY 2148 gaagtggagaggttaccagaagaagttagaagggttaccgcagagaagttagaagaagtggaa 2207
Db 19831 GCAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19772
QY 2208 gaggtaccagaagaagttagaagggttaccagaagaagtggagaggttaccagaagaagttag 2267
Db 19771 GCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19712
QY 2268 gaagggttaccagaagaagttagaagaagtggagaggttagaagaagttagaggttaccagcg 2327
Db 19711 GGAGGAGCAGGAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19652
QY 2328 gtagtagaagttagaagttagcagcggttagtagaagaagaagtgccccagaagaagttagaaga 2387
Db 19651 GTTAGAGGAGCAGGAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCAGGATGAGCAGCA 19592
QY 2388 gaagaagaagagaagaaccagtagagggagaagaagttagtaacatttagtaataaccatcg 2447
Db 19591 GGAGTTAGGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCAGGATGAGCAGCAGCA 19532
QY 2448 gaagaagatatacaattag 2466
Db 19531 GCAGGAGTTAGAGGAGCAG 19513

```

```

RESULT 8
5231168-1
; Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN; BORRE, MARTIN; JEPSEN, SOREN;
; VUUST, JENS; RIENECK, KLAUS; WIND, ANNETTE; JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO: 1
; LENGTH: 3095
5231168-1

```

Query Match 5.2%; Score 175; DB 6; Length 3095;
 Best Local Similarity 44.9%; Pred. No. 4, 9e-23;
 Matches 848; Conservative 0; Mismatches 1020; Indels 21; Gaps 4;

```

QY 597 gaaactggcgatttggaaagcaggggaaagaaactggcgatttggaaagcaggggaaagagct 656
Db 16 gaaaaaacgagcatgaagcagctgaaatgaagaaagttagcttgaagaagggccatcat 75
QY 657 ggtgatttggaaagcaggggaaagaaactggcgatttggaaagcaggggaaagaaactggagat 716
Db 76 gaagaattgtacctgaacaaataatgaagaatcgagtgaaagtgaatttagttgataat 135
QY 717 ggggaaactgaagaagagcaactggagatcgcgaaactgaaatggagcaactgtgtat 776
Db 136 gatgaagggtgttttgaagaagctcatgaaaaatttttcactcctgaagtaagtaactct 195
QY 777 gtgacacagaagatagttcagctgatggagcagaaaaagatacatgttctctgctcaagaa 836
Db 196 gaattaaatgaaaaatgaatttggatctgcacaaagtgtaaactgaacctgctgaacat 255
QY 837 aatgtacaacctgccgatagtaatgatgccctcttggaaagtatttggataaagatatata 896
Db 256 gaagaa-----gtgtatctgaagaagcaacctgaacctgaacctgaagaagt 309
QY 897 atttttgatcataatgaagatttcgagccactattcgacaacaaatttggcggtactgct 956
Db 310 agtatgaagaagctcatcaggaaagaattgtacctggaacaaatgatgaagaatcaggt 369
QY 957 aaactgttacgggacaagaattgccaatgaaacctgtaccattaccagtggcagaagag 1016
Db 370 gaaagtgtattagttgataatgaagaaggtgatttgaagaacctaatcatgaagaattt 429
QY 1017 ccgcgcagtagtaccagcggaagaattagatgccactccagagagtagacttcgacttagat 1076
Db 430 gaacctgtacaaaatgactctgaattgaagtgaataatgaatttagtgaatcagaaaaagt 489
QY 1077 gttacagaatctcccgagggaagttagaatttagtattagtagaaggagcaactggagagaa 1136
Db 490 gtatctgaaccagctgaacatgtagaattgtatcgaaaaaaagtgtatctgaaccagct 549
QY 1137 tcaocggaggtgggaccacgggaagagacacccagacgaagaatttagatgccatccagag 1196
Db 550 gaacacgtagaattgtatctgaaaaaagtacatccgaa-----ccagctgaaca 599
QY 1197 gatggtatttcgattagacgaactcgagaagagagaacagagaacgttagagggagaa 1256
Db 600 tgtagaagtgtatctcgacaaagtataacgaacctccga--aaagaaagatggacca 657
QY 1257 gaacacagaagaactgcagaagagagaagtatcagaagaaactccagaagagagaagagag 1316
Db 658 gttccttcaaacatttgaagaaattgaaaaagtgtgttcaacctaaaaattgtagac 717
QY 1317 tttagaggaactccagagagatgatttcgacttagatggaactacattagaagaaccgaa 1376
Db 718 cttcaataattgaacctaattttgttactcacacaaactccacaaagaccagttgaa 777
QY 1377 gaaactgcagaagagagaacccgtagaggggagaagaacccgtagagggagagaagaacc 1436
Db 778 ccatcatttgcataattgaaaaagtctctctgaagaaataataacatgcaagtgttgat 837
QY 1437 gttagagggagaagaactgcagaagagagaagaagttagaggaactccagagggatgac 1496
Db 838 cctgaagtaaaagaaaaaagaaatgtatctgaagtgttgaagaaaaacaaattccaca 897
QY 1497 ttccaattagaagaacacatcaggaagaggaaggggagaggaaggaaggaaggaagga 1556
Db 898 gaatcagttgaagaataatccagtaaatgaggtgaatttgaagatgttccactgaacaa 957
QY 1557 gaaggagaagcgttagtagcagtgccagtagtgccggaacccgttagaagtagtagtacct 1616
Db 958 tttagatttagatcataaaacagttgatccagaataatagtagaagttgaagaataatctca 1017

```



```
Db 1077 agctccaagtggtgtagaaagtgtggtcccaagtggtggaagaagtgtgtagaagaataatgt 1136
Qy 1247 agagggaagaacacagaagaagctgcagaagagagaagtatcagaagaataatcccgagaagg 1306
Db 1137 tgaagaagtgtagctgaaatgttgaagaagtgtagctgaaatgttgaagaagtggt 1196
Qy 1307 agagaagaagtgtagggcaactccagagagtgatttcctattagatggaactcacattaga 1366
Db 1197 agctgaaatgttgaagaagtgtagctgaaatgttgaagaagtgtagctgaaaaatgt 1256
Qy 1367 agaaacccgaagaactgcagaagggagaaacccgtagaggagagaacccgtagaggg 1426
Db 1257 tgaagaatacgttagctccaactgttgaagaatacgttagctccaactgttgaagaataatgt 1316
Qy 1427 agagaacacccgtagagggagaagaactgtagagaaggaagaagaagttagaggcaactcc 1486
Db 1317 agctccaagtggttagaaagtgtgtgctccaagtgttgaagaagtgttagaagaataatgt 1376
Qy 1487 agaggtgactctccaattagaagaaccatcaggaagacaggaaggggaagggagaaggaga 1546
Db 1377 tgaagaagtgtagctgaaatgttgaagaagtgtagctgaaatgttgaagaagtggt 1436
Qy 1547 aggggaagagagaaggaagcgtttagtagcagtgccagtagtggccgaacccgtgagaagt 1606
Db 1437 agctgaaatgttgaagaagtgtagctgaaatgttgaagaagtgtagctgaaaaatgt 1496
Qy 1607 ---agtgactcctgctcagcctgtcaaaccaatgtgctccaaagcagatgaaacttt 1663
Db 1497 tgaagaagtgtagctgaaatgttgaagaagtgtagctgaaatgttgaagaatcgt 1556
Qy 1664 attcgttatattcttagataaacgatttaacgttatcgacagcattacatcctttgagccatt 1723
Db 1557 agctccaactgttgaagaactgtagctccaactgttgaagaatgttagctccaagtggt 1616
Qy 1724 atttaacaaatccctcaagggaacctgtagcaggaagggctgttaacagtagtaccatcaaaaga 1783
Db 1617 ttagaagaagtgtggtcccaactgttgaagaagtgtagaagaataatgttgaagaagtggt 1676
Qy 1784 agcacctgtacaagtaccagtgagcagtagggcccgcaagaagtgccaacggaagaatt 1843
Db 1677 agctgaaatgttgaagaagtgtagctgaaatgttgaagaagtgtagctgaaaaatgt 1736
Qy 1844 gatgcaactccaagggacgatttcgaattagaagaactgcagaagctccagaggaagg 1903
Db 1737 tgaagaagtgtagctgaaatgttgaagaactgttagctccaactgttgaagaactgt 1796
Qy 1904 agaatttagtattagaaggaaggggaacccacggaagaagggccaaaggaaggagagcc 1963
Db 1797 agctccaactgttgaagaataatgttagctccaactgtttaga-----aagtggtgctcc 1850
Qy 1964 aacagaagagagaagtgccagaagaagaattagaggcaactccagagcgagcagatttcgaatt 2023
Db 1851 agtggtgaagaagtgtagaaggaataatgttgaagaagtgtagctgaaaaatgttgaaga 1910
Qy 2024 agagaacacacagagaagaagtagtaagaacccgttagggcggaagaactgcagaagg 2083
Db 1911 agtgtagctgaaatgttgaagaagtgtagctgaaatgttgaagaactcgttagctcc 1970
Qy 2084 agagaagtggaagaggtaccctgcagaaagttagaagaagtgggaagaggtaccctgcagaagt 2143
Db 1971 aactgttgaagaatacgttagctccaactgttgaagaataatgttagctccaactgtttaga 2030
Qy 2144 agagaagtggaaggttaccagaagaagtagaagaaggtaccgcgagaagtagagaagt 2203
Db 2031 agtggtggtcccaagtgttgaagaagtgtagaagaataatgttgaagaagtgtagctga 2090
Qy 2204 ggaagaggtaccagaagaagtggaagaggtaccagaagaagtaggaagaggtaccagaaga 2263
Db 2091 aaatgttgaagaagtgtagctgaaatgttgaagaagtgtagctgaaaaatgttgaaga 2150
Qy 2264 agtggaagaggttaccagaagaagtggaagaagtggaagaagtagaagaagtagaggtacc 2323
```

```
Db 2151 aagtgtagctgaaaaatgttgaagaatacgttagctccaactgttgaagaatacgttagctcc 2210
Qy 2324 aqcgtagtagaagtagaagtagtaccagcggttagtagaagaagaaggtgtccagaagaagtaga 2383
Db 2211 aactgttgaagaataatgttagctccaagtggttagaagaagtggtggtcccaagtggtgaaga 2270
Qy 2384 aqaagaagaagaaggggaagaccagtagaggaagaagatgtattacaatttagtaatacc 2443
Db 2271 aagtgtagaagaataatgttgaagaagtgtagctgaaaaatgttgaagaagaagtgtagctga 2330
Qy 2444 atcgagaagaagatcacataattagacaa 2470
Db 2331 aatgttgaagaagtgtagctgaaaa 2357

RESULT 10
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1
```

Query Match 5.1%; Score 170.2; DB 4; Length 6152;
Best Local Similarity 44.1%; Pred. No. 4.1e-22;
Matches 911; Conservative 0; Mismatches 1138; Indels 18; Gaps 4;

```
Qy 407 tgaggaaagtcacgaaacgttagatggaggtgaatcattaggagaggtttccactgaaga 466
Db 550 tgaagaagtaaaaggaaaaattcttgactattagaagaaggaatacacattaaactgaag 609
Qy 467 tatggaacacagaagatggtcacaagatgatacggaaaaacagaagaaggaaggtacctggtga 526
Db 610 tgtagatgataataaaaaatttagaagaagccgaagataaaaggaaaaatatcttataag 669
Qy 527 tatggaaggaagaagaagcgtggcgatattggaagcagaggggaagaactggcgatttggaaagcag 586
Db 670 taatacagaagaaccaaagtgatcacgaaaaatattatgcacatttattataatttggacaaaa 729
Qy 587 agcagggggaagaactggcgatttggaaagcaggggaagaactggcgatttggaaagcag 646
Db 730 ttcagaaaaaacagaagaagtgatcacgaaaaatgcacagtcagtgatgaactttttaaaga 789
Qy 647 ggaagaagctggtgatttggaaagcaggggaagaactggcgatttggaaagcaggggaaga 706
Db 790 attattaaatagtgtagtgtaattggaagaagtaaaagaaaaatattttggaggaaagtca 849
Qy 707 aactggagtcggaaaaactgaagaaggaagcaactggagatgcggaactgaaaaatggagc 766
Db 850 agttaatgcgataatttttaattagtttagtaaaaagtgttcaacaagaacaaacacaa 909
Qy 767 aactgtgtgtagcacacagaagatagttcagctgtagtgagcagaaaaagtagcttcc 826
Db 910 tgttgaagaaaaagttgaagaagtgtagaagaataatgacgaagaagtgtagaagaaaa 969
Qy 827 tgcctgaagaaaaatgtacaacctcccgatagtaatgatgcctcttttggaaagtagtttttgg 886
```



```

; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;
; NUMBER OF SEQUENCES: 21
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Genentech, Inc.
;
; STREET: 460 Point San Bruno Blvd
;
; CITY: South San Francisco
;
; STATE: California
;
; COUNTRY: USA
;
; ZIP: 94080
;
; COMPUTER READABLE FORM:
;
; MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
;
; COMPUTER: IBM PC compatible
;
; OPERATING SYSTEM: PC-DOS/MS-DOS
;
; SOFTWARE: patin (Genentech)
;
; CURRENT APPLICATION DATA:
;
; APPLICATION NUMBER: US/08/194.087
;
; FILING DATE: 18-MAY-1992
;
; CLASSIFICATION: 435
;
; PRIOR APPLICATION DATA:
;
; APPLICATION NUMBER:
;
; FILING DATE:
;
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Dreger, Ginger R.
;
; REGISTRATION NUMBER: 33,055
;
; REFERENCE/DOCKET NUMBER: 779
;
; TELECOMMUNICATION INFORMATION:
;
; TELEPHONE: 415/225-3216
;
; TELEFAX: 415/952-9881
;
; TELEX: 910/371-7168
;
; INFORMATION FOR SEQ ID NO: 15:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 10596 bases
;
; TYPE: nucleic acid
;
; STRANDEDNESS: single
;
; TOPOLOGY: linear
;
; US-08-194-087-15

```

Query Match	5.0%	Score 167.2;	DB 2;	Length 10596;
Best Local Similarity	52.1%	Pred. No. 1.6e-21;		
Matches 373;	Conservative	0;	Mismatches 343;	Indels 0;
				Gaps 0;
QY	1739	caagatccctgatgcagagagcgctgaacagtaccatcaaaaggaagcacctgtacaagt	1798	
DB	2209	CAAAAGGACCCACGGTGGACAGGACGAGGACGAGGACGGGGCGCAGGACGAGG	2268	
QY	1799	accagtgcagtaggccccgcgaagaagtgcacaacggaagaattgatgcacctccaaga	1858	
DB	2269	GGCAGGAGCAGGAGGGGCGAGCAGCAGGAGGAGGGCGAGGGCGAGGGGCGAGG	2328	
QY	1859	ggacgatttcgaattagaaggaactgcagaagctccagaggaagagagaattagttattaga	1918	
DB	2329	AGGGGCAGGAGCAGGAGGAGGGCGAGCAGCAGGAGGAGGGCGAGGGCGAGGGGCG	2388	
QY	1919	agggaagggagaaccacgdaagaagccaagaagagagagagagagccacagagaaggaagt	1978	
DB	2389	AGGAGCAGGAGGAGGGCGAGGAGCAGGAGGAGGAGGGCGAGGAGGCGAGGAGGAGG	2448	
QY	1979	gccagaagaagaattagagccaactccagagagacgatttcgaattagaagaaccaacagg	2038	
DB	2449	GGCAGGAGGGGCGAGGAGGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGAGGGGCGAGG	2508	
QY	2039	agaagaagttagaagaaccgttagaggccgaagaactgcagaagaagagaagaagtggaaag	2098	
DB	2509	AGGGGCAGGAGCAGGAGGAGGGGCGAGGAGGGCGAGGAGGGCGAGGAGGAGGAGGGCG	2568	
QY	2099	ggtacctgcagaagttagaagaagtggaaagaggttaccctgcagaagtagaagaagtggaaaga	2158	
DB	2569	AGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGAGGAGG	2628	
QY	2159	ggtaccagaagaadtagaagaaggttaccocgagaagtagaagaagtggaagaggtaccaga	2218	
DB	2629	AGGGGCAGGAGGGGCGAGGAGGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGAGGGGCGAGG	2688	

Qy	2219	agaagtggaaaggtgaccagaagaagtggaaaggtgaccagaagaagtggaaaggtgacc	2278
Db	2689	AGGAGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGG	2748
Qy	2279	agaagaagtggaaaggtgaaagaagtgaagaagtgaagaagtgaagaagtgaagaagtgaagaagt	2338
Db	2749	AGCAGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGG	2808
Qy	2339	agaagtaccagcggtgtagtgaagaagaggtgccagaagaagtgaagaagaagtgaagaagaaga	2398
Db	2809	AGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGG	2868
Qy	2399	ggaagaaccagttagagaagaagtgtattacaattagtaatacacaatcggaagaag	2454
Db	2869	GGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAG	2924
RESULT 20			
PCT-US93-04648-15			
; Sequence 15, Application PC/TUS9304648			
; GENERAL INFORMATION:			
; APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A.,			
; TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS			
; NUMBER OF SEQUENCES: 21			
; CORRESPONDENCE ADDRESS:			
; ADDRESSEE: Genentech, Inc.			
; STREET: 460 Point San Bruno Blvd			
; CITY: South San Francisco			
; STATE: California			
; COUNTRY: USA			
; ZIP: 94080			
; COMPUTER READABLE FORM:			
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk			
; COMPUTER: IBM PC compatible			
; OPERATING SYSTEM: PC-DOS/MS-DOS			
; SOFTWARE: patin (Genentech)			
; CURRENT APPLICATION DATA:			
; APPLICATION NUMBER: PCT/US93/04648			
; FILING DATE: 19930517			
; CLASSIFICATION:			
; PRIORITY APPLICATION DATA:			
; APPLICATION NUMBER: 07/884811			
; FILING DATE: 18-MAY-92			
; PRIORITY APPLICATION DATA:			
; APPLICATION NUMBER: 07/885971			
; FILING DATE: 18-MAY-92			
; ATTORNEY/AGENT INFORMATION:			
; NAME: Dreger, Ginger R.			
; REGISTRATION NUMBER: 33,055			
; REFERENCE/DOCKET NUMBER: 755,779P1			
; TELECOMMUNICATION INFORMATION:			
; TELEPHONE: 415/225-3216			
; TELEFAX: 415/952-9881			
; TELEX: 910/371-7168			
; INFORMATION FOR SEQ ID NO: 15:			
; SEQUENCE CHARACTERISTICS:			
; LENGTH: 10596 bases			
; TYPE: NUCLEIC ACID			
; STRANDEDNESS: single			
; TOPOLOGY: linear			
; PCT-US93-04648-15			

	Query Match	5.08;	Score 167.2;	DB 5;	Length 10596;
	Best Local Similarity	52.1%;	Pred. No. 1.6e-21;		
	Matches 373;	Conservative	0;	Mismatches 343;	Indels 0; Gaps 0;
QY	1739	caagatcctgatgcagagagaggtgttaacagtgaccatcaagaagaagcacctgtacaagt	1798		
Db	2209	CAAAAGGACCCACCGTGTGAACAGGAGCAGGAGCAGGAGCGGAGGGGCGAGGACGAGG	2268		
QY	1799	accagtggcagtagggccgcgcgaagaagtgtccaacggaagaattgtatgcactccaaga	1858		

Db 2269 GGCAGGAGCAGGAGGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGGCGAGG 2328
Qy 1859 ggcagatttcgaattagaagaactcagaagctccagagaaggaaggaattagattaga 1918
Db 2329 AGGGCGAGGACGAGGAGGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGGGCG 2388
Qy 1919 agagaaggagaaacacggaagaagcgaagaggaagcgaagaggaaggaaggaag 1978
Db 2389 AGGAGCAGGAGGAGGGGCGAGGAGCAGGAGGAGGGGCGAGGAGGGGCGAGG 2448
Qy 1979 gcagaagaagaattagaggcaactccagaggacgatttcgaatttagaagaacacacag 2038
Db 2449 GGCAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGG 2508
Qy 2039 agaagaagttagaagaacacctagagggcggaagaactcgaagaggaaggaagaaga 2098
Db 2509 AGGGCGAGGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCG 2568
Qy 2099 ggtacctgcagaagttagaagaagtgggaaggttacctgcagaagttagaagaagtgg 2158
Db 2569 AGGAGCAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2628
Qy 2159 ggtaccagaagaagttagaaggttaccgcgagaagtagaagaagtgggaaggtaccaga 2218
Db 2629 AGGGCGAGGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 2688
Qy 2219 agaagttagaaggttaccagaagaagtgggaaggttaccagaagaagtgggaaggtacc 2278
Db 2689 AGGAGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGG 2748
Qy 2279 agaagaagttagaagaagttagaagaagttagaagaagttagaagaagttagaagaagt 2338
Db 2749 AGCAGGAGGGGCGAGGAGGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGG 2808
Qy 2339 agaagtaccagcgttagaagaagaaggtgccagaagaagttagaagaagaagaagaaga 2398
Db 2809 AGGGCGAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGG 2868
Qy 2399 ggaagaaccagtagagaagaagaagttagaagaagttagaagaagttagaagaagaag 2454
Db 2869 GGCAGGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGGCGAGGAGGGG 2924

RESULT 21
US-08-056-200-93
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Methods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: Fedrick, Michael F.
REGISTRATION NUMBER: 36,799
REFERENCE/DOCKET NUMBER: NIH054.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (714) 760-0404
TELEFAX: (714) 760-9502
INFORMATION FOR SEQ ID NO: 93:
SEQUENCE CHARACTERISTICS:
LENGTH: 9551 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 1507..1644
FEATURE:
NAME/KEY: Intron
LOCATION: 1645..2511
FEATURE:
NAME/KEY: CDS
LOCATION: 2512..8070
US-08-056-200-93

Query Match 4.4%; Score 147.2; DB 1; Length 9551;
Best Local Similarity 47.2%; Pred. NO. 5.9e-18;
Matches 566; Conservative 0; Mismatches 613; Indels 21; Gaps 3;

Qy 1231 gaacacagaagaacgttagagggaagaagaagctgcagaaggaaggaagtcacg 1290
Db 3190 GAGGAAGAAGCAGCAGCTACCGAAGCTGAGCGCGCAGAGCTGAGGAGGAGCGCCAGGAG 3249
Qy 1291 aagaactccagaaggaagaaggttagaggcaactccagaggtatgttcgcattag 1350
Db 3250 GAGAGCAGCAGCAGCAAGAGCTGAGCGCGCAGCAGCACTAAGCGCAAGCAGAGGAGG 3309
Qy 1351 atggaactacatagaagaacccaagaactcagaaggaagaagaaccgttagagagag 1410
Db 3310 GAGAGGCGCGCAGCAGGAGGAGCGCGCAGCAGGAGCGCGCAGCAGGAGCGCGCAGCAGGAG 3369
Qy 1411 aagaacacgttagagggaagaagaacacctgtagagggaagaagctgcagaaggaagaag 1470
Db 3370 GAGAGGCGCGCAGCAGCAGCTGAGCGCGCAGCAGGAGGAGGAGCGCGCAGCAGCAGCTGAGG 3429
Qy 1471 agttagaggcaactccagaggttagcttccaatttagaagaacatccatcagggaaggaag 1530
Db 3430 CGCGCAGCAGGAGGAGGAGCGCGCAGCAGCAGCTGAGGCGCGCAGCAGGAGGAGGAGGAGG 3489
Qy 1531 ggaagaagaagaagaaggggaagaagaagaagcgttagtagcagtgccagtagtg 1590
Db 3490 CGCGCAGCAGCAGCTGAGCGCGCAGCAGGAGGAGGAGGAGCGCGCAGCAGCAGCTGAGGCGCG 3549
Qy 1591 ccgaacccgttagaagtagtgactctctgctcagcctgtccaaccaatggtcgctccacag 1650
Db 3550 GAGCAGCAGCTGAGGGCGCAGCAGCAGCTGAGGGCGCGCAGCAGCAGCTGAGGGCGCGCAGCAG 3609
Qy 1651 cagatgaactttatttggttatcttagataacgatttaacgtatcgagacattacat 1710
Db 3610 CAGCTGAGGCGCGCAGCAGCAGCT---GAGGCGCGCAGCAGCAGCTGAGGCGCGCAGCAGCAG 3666
Qy 1711 ccittgagccattatttaacaaatcctcaagatcctgagcaggaaggaagcgtgtaacag 1770
Db 3667 CTGAGGCGCGCAGCAGCAGCTGAGCGCGCAGCAGCAGCTGAGGCGCGCAGGAGGAGGAGGAG 3726
Qy 1771 taccataaaggaagccctgtacaagtaccagtggcagtagggcccgccgcaagaagtgc 1830
Db 3727 AGGCACGAGCAGAGCAGCAGCAGGAGGAGGCGCGCGCTGAGGCGCGCAGCAGGAGGAG 3786
Qy 1831 caacgggaagaattgatgtaactccaagaggaagcagatttcga---attagaaggaagtcag 1887


```
QY 1940 agaagaccgaagaagagagccacaagaagagagagtagtccagagaagaattagggc 1999
Db 168 atgtccacagaaaaaggagagaaaaagaagaagcagtgccagcagaataaaaaatga 227
QY 2000 aactccagagcagattctcaattagaagaaccacagcaggaagaagtagaagaaccgct 2059
Db 228 agaagaagatcagaagaagatgaagaagatcaaaacgaagagaagaagggaagctggaaa 287
QY 2060 agagggcgaaacactgcagaagagaagaagtggaaaggtaccctgcagaagtagaaga 2119
Db 288 agaagacaagatgaaaaagggaagaagatcggaagaagagataaaaaatggaaatgagaa 347
QY 2120 agtggaaaggtaccctgcagaagtagaagaagtggaaaggtaccagagaagaagtagaaga 2179
Db 348 aggaagaagtgcacaaagagaagaagatgaaaaaaagtgaagacggaaaaaggaaatgg 407
QY 2180 ggtaccgcagaagtagaagaagtggaaaggtaccagaagaagtggaagaggtaccaga 2239
Db 408 agaagatggaaaagagaagaagatgaaaaaggggaagaagacagaaaaagaaacagg 467
QY 2240 agaagtggaagaggtaccagaagaagtggaaaggtaccagaagaagtggaaagtgga 2299
Db 468 agttgaaaagaagaatgaagatggaaaagagaaggagagataaaaaagagggggaaagatgt 527
QY 2300 agaagtagaagaagtagaggtaccagcggtagtagaagtagaagtagtaccagcggtagtaga 2359
Db 528 aaaagtcacaaagaagtagaaaagagagagaagatggaaaagaagatgaagtggaatga 587
QY 2360 agaagaaggtccagaagaagtagaagaagaagaagaagaagaagacacagtagagaaga 2419
Db 588 ggaagaagctggaaaagagaagaagatttaaaagaagagagaagggaagaagggaaga 647
QY 2420 agatgtattacaa 2432
Db 648 tgagatcaaaaga 660

RESULT 25
US-09-461-697-189
; Sequence 189, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 189
; LENGTH: 717
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-189
```

```
Query Match 4.3%; Score 145; DB 4; Length 717;
Best Local Similarity 53.9%; Pred. No. 8.2e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 1880 aactgcagaagctccagagaagagaattagtagtagaaggagaaggaaccaacgga 1939
Db 126 aattacagaggccacagcttctgaaaaagaattgtggaagtagtaaaagaagaataattga 185
QY 1940 agaagaccgaagaagagagcacaagaagagagaagtagtccagagaagaattagaggc 1999
```

```
Db 186 atgtgccacagaaaaaggagagaaaaagaagaagcagtgccagcagaataaaaaatga 245
QY 2000 aactccagagcagattctgaattagaagaaccacagcaggaagaagtagaagaaccgct 2059
Db 246 agaagaagatcagaagaagatgaagaagatcaaaacgaagagaagaagggaagctggaaa 305
QY 2060 agagggcgaaacactgcagaagagaagaagtggaaaggtaccctgcagaagtagaaga 2119
Db 306 agaagacaagatgaaaaagggggaagaagatggaaaagaggataaaaaatggaaatgagaa 365
QY 2120 agtggaaaggtaccctgcagaagtagaagaagtggaaaggtaccagaagaagtagaaga 2179
Db 366 aggaagaagatgcaaaagagaagaagatggaaaaaaggtagaagacggaaaaagaaatgg 425
QY 2180 ggtaccgcagaagtagaagaagtggaaaggtaccagaagaagtggaagaggtaccaga 2239
Db 426 agaagatggaaaagagaagaagatgaaaaaggggaagaatgaaaaaggggaagacagaaaaacagg 485
QY 2240 agaagtggaagaggtaccagaagaagtggaaaggtaccagaagaagtggaaagtgga 2299
Db 486 agttgaaaagaagaatgaagatggaaaagagaaggagagataaaaaagagggggaaagatgt 545
QY 2300 agaagtagaagaagtagaggtaccagcggtagtagaagtagaagtagtaccagcggtagtaga 2359
Db 546 aaaagtcacaaagaagtagaaaagagagagaagatggaaaagaagatgaagtggaatga 605
QY 2360 agaagaaggtgccagaagaagtagaagaagaagaagaagaagaggaagacacagtagagaaga 2419
Db 606 ggaagaagctggaaaagagaagaagatttaaaagaagagagaagggaagaagggaaga 665
QY 2420 agatgtattacaa 2432
Db 666 tgagatcaaaaga 678
```

```
RESULT 26
US-09-461-697-187
; Sequence 187, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 774
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-187
```

```
Query Match 4.3%; Score 145; DB 4; Length 774;
Best Local Similarity 53.9%; Pred. No. 8.3e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;

QY 1880 aactgcagaagctccagagaagagaattagtagtagaaggagaaggaaccaacgga 1939
Db 183 aattacagaggccacagcttctgaaaaagaattgtggaagtagtaaaagaagaataattga 242
QY 1940 agaagaccgaagaagagagcacaagaagagagaagtagtccagagaagaattagaggc 1999
```

Db 243 agatgccacagaagaggaggagaaagaaagcagtgccagcagaagtataaaatga 302
QY 2000 aactccagaggacgatttcgaattagaagaaccacacagagagaagtagaagaacccgt 2059
Db 303 agaagaagatcagaagaagatgaagaagatcaaacgaagagaaggggaagctggaaa 362
QY 2060 agaggccgaagaactgcagaaagagaagaagtgcgaagaggtaccctgcagaagttagaaga 2119
Db 363 agaagcaaaagatgaaaaaggggaagaagatggaagagaggataaaaaatggaaatgaga 422
QY 2120 agtgggaaggtaccctgcagaagttagaagaagtgcgaagaggtaccagaagaagttagaaga 2179
Db 423 agagaagaatgcgaagagagaagaagatggaagaaaggtggaagacggaaggaatgg 482
QY 2180 ggtaccgcgagaaagtgaagaagtggaagagggtaccagaagaagtggaagaggtaccaga 2239
Db 483 agaagatggaagaagaaaggagaagatgaaaagaggaagacagacagaaaagaaacagg 542
QY 2240 agaagtgcgaagaggtaccagaagaagtgcgaagagggtaccagaagaagtgcgaagaagtgcga 2299
Db 543 agtggaaaagagaatgaagatggaagagagaagggagataaaaaagaggggaaagatgt 602
QY 2300 agaagtagaagaagttagaggtaccagcggttagtagaagttagaagtagaccagcggttaga 2359
Db 603 aaagtcacaagaagatgaaaagagagagaagatggaagaaagatggaaggtggaatga 662
QY 2360 agaagaggtgccagaagaagttagaagaagaagaagaagagagaagaccagtagaggaaga 2419
Db 663 ggaagaagctggaagaagaagaagatttaaaagagaggaaggaaggaaggaagaaga 722
QY 2420 agatgtattacaa 2432
Db 723 tgagatcaaaagaa 735

RESULT 27
US-09-461-697-185
; Sequence 185, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 185
; LENGTH: 819
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-185

Query Match 4.3%; Score 145; DB 4; Length 819;
Best Local Similarity 53.9%; Pred. No. 8.4e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;
QY 1880 aactgcagaagctccagagaagagaattagtagtagaaggagaaggaacacagga 1939
Db 228 aattacagagccaccagctcttgaagaaagaaattgtggaagtaaaagagaataattga 287
QY 1940 agaagaccaagaagagagagccacagagaagagaagtagtccacagaagaagaattagaggc 1999
Db 288 agatgccacagaagaggaggaagaaagaaagcagtcagtcgagcagaagataaaaaatga 347

QY 2000 aactccagaggacgatttcgaattagaagaaccacacagagagaagtagaagaacccgt 2059
Db 348 agaagaagatcagaagaagatgaagaagatcaaacgaagagaaggggaagctggaaa 407
QY 2060 agaggccgaagaactgcagaaagagaagaagtgcgaagaggtaccctgcagaagttagaaga 2119
Db 408 agaagcaaaagatgaaaaaggggaagaagatggaagagaggataaaaaatggaaatgaga 467
QY 2120 agtgggaaggtaccctgcagaadttagaagaagtgcgaagaggtaccagaagaagttagaaga 2179
Db 468 agagaagaatgcgaagagagaagaagatggaagaaaggtggaagacggaaggaatgg 527
QY 2180 ggtaccgcgagaaagtgaagaagtggaagagggtaccagaagaagtggaagaggtaccaga 2239
Db 528 agaagatggaagaagagaagaggagaagatgaaaagaggaagacagacagaaaagaaacagg 587
QY 2240 agaagtgcgaagaggtaccagaagaagtgcgaagagggtaccagaagaagtgcgaagaagtgcga 2299
Db 588 agtggaaaagagaatgaagatggaagagagaagggagataaaaaagaggggaaagatgt 647
QY 2300 agaagtagaagaagttagaggtaccagcggttagtagaagttagaagtagaccagcggttaga 2359
Db 648 aaagtcacaagaagatgaaaagagagagaagatggaagaaagatggaaggtggaatga 707
QY 2360 agaagaggtgccagaagaagttagaagaagaagaagaagagagaagaccagtagaggaaga 2419
Db 708 ggaagaagctggaagaagaagaagatttaaaagagaggaaggaaggaaggaagaagaaga 767
QY 2420 agatgtattacaa 2432
Db 768 tgagatcaaaagaa 780

RESULT 28
US-09-461-697-184
; Sequence 184, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Puranam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 184
; LENGTH: 1669
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-461-697-184

Query Match 4.3%; Score 145; DB 4; Length 1669;
Best Local Similarity 53.9%; Pred. No. 9.9e-18;
Matches 298; Conservative 0; Mismatches 255; Indels 0; Gaps 0;
QY 1880 aactgcagaagctccagagaagagaattagtagtagaaggagaaggaacacacagga 1939
Db 307 aattacagagccaccagctcttgaagaaagaaattgtggaagtaaaagagaataattga 366
QY 1940 agaagaccaagaagagagagccacagagaagagaagtagtccacagaagaagaattagaggc 1999
Db 367 agatgccacagaagaggaggaagaaagaaagcagtcagtcgagcagaagataaaaaatga 426

Search completed: June 14, 2002, 15:29:39
Job time: 19070 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 11:20:08 ; Search time 350.64 Seconds
(without alignments)
16339.675 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagaagtaacaa.....tcttataaaataaataatc 3337

Scoring table: IDENTITY_NUC
Gapex 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	17 AAT34620	P. vivax ESP-1 blo
2	3337	100.0	3337	20 AX15174	DNA encoding a sec
3	3337	100.0	3337	22 AAH76457	Plasmodium vivax E
4	342.8	10.3	3399	17 AAT05868	Chicken leucocytoz
5	268	8.0	1686	16 AAQ87587	DNA encoding Leuco
6	228	6.8	3579	21 AAA70099	Plasmodium falcipa
7	220.2	6.6	2805	22 AD07130	Canine retinitis p
8	219	6.6	2019	24 ABA90848	Bacillus anthracis
9	217.4	6.5	2849	22 AAD21684	Human retinitis pi

10	214.6	6.4	2014	24 ABA90791	Bacillus anthracis
11	211.6	6.3	2800	22 AAD07131	Canine retinitis p
12	211	6.3	2803	22 AAD07132	Canine retinitis p
13	200.6	6.0	1050	22 AAD21685	Mutational hot spo
14	199.6	6.0	621	23 AAS90688	DNA encoding novel
15	199	6.0	3489	21 AAA30290	Kaposi's sarcoma-a
16	199	6.0	3489	22 AAF82901	Nucleotide sequenc
17	199	6.0	32207	20 AAV73805	KSHV LUR DNA (nucl
18	199	6.0	137507	19 AAV19941	KSHV long unique c
19	181.8	5.4	4954	23 ABL05187	Drosophila melanog
20	174	5.2	9516	23 ABL05186	Drosophila melanog
21	171.8	5.1	3095	11 AAQ03875	Sequence encoding
22	170.2	5.1	5361	18 AAT78868	P. falciparum live
23	170.2	5.1	6152	18 AAT78867	P. falciparum live
24	169.4	5.1	49999	20 AAZ23891	Murine LOBO genom
25	169.4	5.1	49999	20 AAZ23896	Murine LOBO genom
26	168.8	5.1	4590	7 AAN60472	Sequence encoding
27	168.4	5.0	585	16 AAQ87589	DNA encoding Leuco
28	168.4	5.0	996	16 AAQ87588	DNA encoding Leuco
29	167.2	5.0	1926	21 AA50254	Epstein Barr virus
30	167.2	5.0	1926	22 AAF82902	EBV tethering p
31	167.2	5.0	2580	21 AA75454	Nucleotide sequenc
32	167.2	5.0	5452	20 AA90923	Anti-sense strand
33	167.2	5.0	8705	20 AAZ23778	Vector pShuttle DN
34	167.2	5.0	9600	19 AAV21683	Vector plasmid PCM
35	167.2	5.0	10380	20 AAZ22248	Nucleotide sequenc
36	167.2	5.0	10596	14 AAQ51731	Plasmid pCISERON f
37	167.2	5.0	10596	17 AAT40348	Plasmid pCISERON f
38	167.2	5.0	10596	20 AX15650	Nucleotide sequenc
39	167.2	5.0	16080	21 AA59553	DNA clone pCEK Cl.
40	165.8	5.0	799	19 AAV55831	Nucleotide sequenc
41	165.8	5.0	1218	23 AAS90730	DNA encoding novel
42	164.4	4.9	1215	23 AAS92075	DNA encoding novel
43	164.2	4.9	3662	23 ABL05274	Drosophila melanog
44	160	4.8	8201	21 AAR88864	Human dentin sialo
45	159.8	4.8	40875	18 AAT80043	Insert from cosmid

ALIGNMENTS

RESULT 1
AAT34620
ID AAT34620 standard; DNA; 3337 BP.
XX
AC AAT34620;
XX
DT 12-NOV-1996 (first entry)
XX
DE P. vivax ESP-1 blood stage antigen coding sequence.
XX
KW ESP-1; blood stage antigen; diagnosis; malaria; infection;
KW causative agent; antibody; monoclonal; polyclonal; assay; ds.
XX
OS Plasmodium vivax (clone PvMB3.3.1).
XX
FH Key Location/Qualifiers
FT Exon 1..91
FT FT /*tag= a
FT FT /note= "encodes initial (N-terminal) sequence of
FT FT intron 92..230
FT FT /*tag= b
FT FT /note= "contains typical malaria intervening
FT FT exon 231..3197
FT FT /*tag= c
XX
US5532133-A.
XX
PD 02-JUL-1996.
XX
PD 02-JUN-1993; 93US-0072610.

QY 1741 aggatcctgatgcaggagaggtgttaacagtagtaccatacaaggagagcacctgtacaagtac 1800
|||||
Db 1741 aggatcctgatgcaggagaggtgttaacagtagtaccatacaaggagagcacctgtacaagtac 1800

QY 1801 cagtggcagtagggccgcgcgaagaagtcgcaacggaagaattgatgcaactcccaagagg 1860
|||||
Db 1801 cagtggcagtagggccgcgcgaagaagtcgcaacggaagaattgatgcaactcccaagagg 1860

QY 1861 acgatttcgaattagaagaactgcagaagctccagaggagaagaattagttattagaag 1920
|||||
Db 1861 acgatttcgaattagaagaactgcagaagctccagaggagaagaattagttattagaag 1920

QY 1921 ggaaggagaaaccaacggaagaagtcgcaacggaagaagccaaacagaaaggagagtgctc 1980
|||||
Db 1921 ggaaggagaaaccaacggaagaagtcgcaacggaagaagccaaacagaaaggagagtgctc 1980

QY 1981 cagaagaagaattagaggaactccagaggaacgatttcgaattagagaaccaaagggag 2040
|||||
Db 1981 cagaagaagaattagaggaactccagaggaacgatttcgaattagagaaccaaagggag 2040

QY 2041 aagaagtagaagaacacgtagaggcggaagaactgcagaagagagaagaagtggaaagg 2100
|||||
Db 2041 aagaagtagaagaacacgtagaggcggaagaactgcagaagagagaagaagtggaaagg 2100

QY 2101 tacctgcagaagttagaagaagtcggaagagg tacctgcagaagttagaagaagtcggaagagg 2160
|||||
Db 2101 tacctgcagaagttagaagaagtcggaagagg tacctgcagaagttagaagaagtcggaagagg 2160

QY 2161 taccgaagaagaatagaagaaggtaccgcgagaagtagaagaagtcggaagagggtaccagaag 2220
|||||
Db 2161 taccgaagaagaatagaagaaggtaccgcgagaagtagaagaagtcggaagagggtaccagaag 2220

QY 2221 aagtggaagagtagcagaagaagtcggaagaggtagcagaagaagtcggaagagggtaccag 2280
|||||
Db 2221 aagtggaagagtagcagaagaagtcggaagaggtagcagaagaagtcggaagagggtaccag 2280

QY 2281 aagaagtggaagaagtgggaagaagtagaagaagtagaggtaccagcggtagtagaagtag 2340
|||||
Db 2281 aagaagtggaagaagtgggaagaagtagaagaagtagaggtaccagcggtagtagaagtag 2340

QY 2341 aagttaccagcggtagtagaagaagaggtgccagaagaagtagaagaagaagaagaagg 2400
|||||
Db 2341 aagttaccagcggtagtagaagaagaggtgccagaagaagtagaagaagaagaagaagg 2400

QY 2401 aagaaccagtagagaagaagatgtattacaattagtaaccatcggaagaagatatac 2460
|||||
Db 2401 aagaaccagtagagaagaagatgtattacaattagtaaccatcggaagaagatatac 2460

QY 2461 aattagacaaaccaaagaagcgaattaggctctggaatttttatctatcatcgacatgc 2520
|||||
Db 2461 aattagacaaaccaaagaagcgaattaggctctggaatttttatctatcatcgacatgc 2520

QY 2521 actacaagaacgttccaaaggattttatggaagaagaagaagaactgcagtgatccat 2580
|||||
Db 2521 actacaagaacgttccaaaggattttatggaagaagaagaagaactgcagtgatccat 2580

QY 2581 tgaaccagaagaattttgcaagggaagattcacaattcagaagaagtcacattcattc 2640
|||||
Db 2581 tgaaccagaagaattttgcaagggaagattcacaattcagaagaagtcacattcattc 2640

QY 2641 aaggcctagaaggcagctgggaacgattagaagtgagcttaataaaggctagagaagaat 2700
|||||
Db 2641 aaggcctagaaggcagctgggaacgattagaagtgagcttaataaaggctagagaagaat 2700

QY 2701 ggaatgaacaaagaaataaagaatggctggctggcttcgcttaattgaaataaattggt 2760
|||||
Db 2701 ggaatgaacaaagaaataaagaatggctggctggcttcgcttaattgaaataaattggt 2760

QY 2761 cagaataatagtcacaaatttcaacaaagggaagaccagctggttttgaaaacgagagt 2820
|||||
Db 2761 cagaataatagtcacaaatttcaacaaagggaagaccagctggttttgaaaacgagagt 2820

QY 2821 ggagcagcagagaaatggaataatggttttaaaagcagaagtcacaaatcccaattgattcac 2880
|||||
Db 2821 ggagcagcagagaaatggaataatggttttaaaagcagaagtcacaaatcccaattgattcac 2880

QY 2881 acttgaataaaatggatgaaacgcacactcattccaaattatttataaattcttctgtgaagata 2940
|||||
Db 2881 acttgaataaaatggatgaaacgcacactcattccaaattatttataaattcttctgtgaagata 2940

QY 2941 tgtcacaaatttgaataaaacaaacaaagaatgggttaattgaatcactggaataaagacg 3000
|||||
Db 2941 tgtcacaaatttgaataaaacaaacaaagaatgggttaattgaatcactggaataaagacg 3000

QY 3001 aacgggggttatgggttctgaaatcatttgaagttatgacacacatcaaaattatttaattgtgg 3060
|||||
Db 3001 aacgggggttatgggttctgaaatcatttgaagttatgacacacatcaaaattatttaattgtgg 3060

QY 3061 ctaagagtcgagaatgggtaccgtgccaaatcctaataataatagagaagaagaactca 3120
|||||
Db 3061 ctaagagtcgagaatgggtaccgtgccaaatcctaataataatagagaagaagaactca 3120

QY 3121 tgaatgggttctcctaaaagaaacgaatattatttagacaaagaatgggaataatggact 3180
|||||
Db 3121 tgaatgggttctcctaaaagaaacgaatattatttagacaaagaatgggaataatggact 3180

QY 3181 cattggaaaaaagttaaaatttttggttcaatttcaattgtaacacattttctggaata 3240
|||||
Db 3181 cattggaaaaaagttaaaatttttggttcaatttcaattgtaacacattttctggaata 3240

QY 3241 cgcttaacaaaggaagaatgggaatcgaattttggttcaattgtaacacattttctggaata 3300
|||||
Db 3241 cgcttaacaaaggaagaatgggaatcgaattttggttcaattgtaacacattttctggaata 3300

QY 3301 aaaaacacagattattctcttataaaaataaataattc 3337
|||||
Db 3301 aaaaacacagattattctcttataaaaataaataattc 3337

RESULT 2
AAK15174
ID AAK15174 standard; DNA; 3337 BP.
XX
AC AAK15174;
XX
DT 28-APR-1999 (first entry)
XX
DE DNA encoding a secreted blood-stage protein called PvESP-1.
XX
KW Erythrocyte secreted protein-1; PvESP-1; malarial antigen;
KW blood-stage protein; malaria; monoclonal antibody ID11G10; ds.
XX
OS Plasmodium vivax.
XX
FH Location/Qualifiers
FT CDS 1..3197
FT /tag= a
FT /note= "contains 1 intron"
FT exon 1..91
FT /tag= b
FT /number= 1
FT intron 92..230
FT /tag= c
FT /number= 1
FT exon 231..3194
FT /tag= d
FT /number= 2
XX
XX US5874527-A.
PN
XX 23-FEB-1999.
PD
XX 30-SEP-1996;
PF 96US-0719822.
XX 02-JUN-1993;
PR 93US-0072610.

Db 1681 ataagatttaacgtatgcagacattacatcctttgagccattatttaacaaaactcctca 1740
Qy 1741 aggatcctgatgcagagagcgctgaacagtagaccatcaagaagaacacacctgtacaaagtac 1800
Db 1741 aagatcctgatgcagagagcgctgttaacagtagaccatcaagaagaacacacctgtacaaagtac 1800
Qy 1801 cagtgagcagtagggcccgccgcaagaagtgcccaacggaagaattgtatgcaactcccaagagg 1860
Db 1801 cagtgagcagtagggcccgccgcaagaagtgcccaacggaagaattgtatgcaactcccaagagg 1860
Qy 1861 acgatttcgaattagaagaacatgcagaagctccagaggaagaagaattagtagtagaag 1920
Db 1861 acgatttcgaattagaagaacatgcagaagctccagaggaagaagaattagtagtagaag 1920
Qy 1921 ggaagaggaacccaacggaagaagcccaagaagaagagcacaacagaagaagaagtagtc 1980
Db 1921 ggaagaggaacccaacggaagaagcccaagaagaagagcacaacagaagaagaagtagtc 1980
Qy 1981 cagaagaagaattagaggcaactccagaggaagcagatttcgaaattagaagaacccaacagggag 2040
Db 1981 cagaagaagaattagaggcaactccagaggaagcagatttcgaaattagaagaacccaacagggag 2040
Qy 2041 aagaagtagaagaacccgtagaggcggaagaacatgcagaagaagaagaagtggaaagagg 2100
Db 2041 aagaagtagaagaacccgtagaggcggaagaacatgcagaagaagaagaagtggaaagagg 2100
Qy 2101 tacctgcagaagtagaagaagtggaagaggtaccctgcagaagtagaagaagtggaagagg 2160
Db 2101 tacctgcagaagtagaagaagtggaagaggtaccctgcagaagtagaagaagtggaagagg 2160
Qy 2161 taccagaagaagtagaagaaggtaccgcgagaagtagaagaagtggaagaggtaccagaag 2220
Db 2161 taccagaagaagtagaagaaggtaccgcgagaagtagaagaagtggaagaggtaccagaag 2220
Qy 2221 aagtgaagagtagaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccag 2280
Db 2221 aagtgaagagtagaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccag 2280
Qy 2281 aagaagtggaagaagtggaagaagtagaagaagtagaggtaccagcggtagtagaagtag 2340
Db 2281 aagaagtggaagaagtggaagaagtagaagaagtagaggtaccagcggtagtagaagtag 2340
Qy 2341 aagtaccagcggtagtagaagaagaaggtgccagaagaagtagaagaagaagaagaagagg 2400
Db 2341 aagtaccagcggtagtagaagaagaaggtgccagaagaagtagaagaagaagaagaagagg 2400
Qy 2401 aagaaccagtagagaagaagatgtattacaattagtaataccatcggaagaagatatac 2460
Db 2401 aagaaccagtagagaagaagatgtattacaattagtaataccatcggaagaagatatac 2460
Qy 2461 aattagacaacccaagaagaacgaattaggctctggaattttatctatcatcgacatgc 2520
Db 2461 aattagacaacccaagaagaacgaattaggctctggaattttatctatcatcgacatgc 2520
Qy 2521 actacaagaacgttccaaagggaatttatggaagaagaagaagaacacgagtgatcccat 2580
Db 2521 actacaagaacgttccaaagggaatttatggaagaagaagaagaacacgagtgatcccat 2580
Qy 2581 tgaaccagaagaattttgcaaaaggaagattccaaactctacagaagaagtcacattcatc 2640
Db 2581 tgaaccagaagaattttgcaaaaggaagattccaaactctacagaagaagtcacattcatc 2640
Qy 2641 aagggctagaagcgagctgggaacgattagaagtgagcttaataaggttagagaagaagat 2700
Db 2641 aagggctagaagcgagctgggaacgattagaagtgagcttaataaggttagagaagaagat 2700
Qy 2701 ggaatgaacaaagaataaagaattggctggctggcttcgcttaattgaaaaataaatgg 2760
Db 2701 ggaatgaacaaagaataaagaattggctggctggcttcgcttaattgaaaaataaatgg 2760
Qy 2761 cagaatatagtcaaaatttcaacaaaagggaagaccacagctgggttgagaagaacagagat 2820
Db 2761 cagaatatagtcaaaatttcaacaaaagggaagaccacagctgggttgagaagaacagagat 2820

Qy 2821 ggagcgacgagaaatgggaaaaaattggttttaaagcagaagtcaaaatccccaaattgattcac 2880
Db 2821 ggagcgacgagaaatgggaaaaaattggttttaaagcagaagtcaaaatccccaaattgattcac 2880
Qy 2881 acttgaaaaaatgggatgaacgacactcatctccaattattttaaattcttggaaagata 2940
Db 2881 acttgaaaaaatgggatgaacgacactcatctccaattattttaaattcttggaaagata 2940
Qy 2941 tctcaaaatttgaacaaagaacaaacgaagaatggttaatgaatcactggaaaaagaacg 3000
Db 2941 tctcaaaatttgaacaaagaacaaacgaagaatggttaatgaatcactggaaaaagaacg 3000
Qy 3001 aacgggggttatggttctgaatcatttgaagttatgaccacatcaaaattattttaaattgtgg 3060
Db 3001 aacgggggttatggttctgaatcatttgaagttatgaccacatcaaaattattttaaattgtgg 3060
Qy 3061 ctaagagtcgagaatgggtaccgtgccaatccctaataataatagagaagaagagaaactca 3120
Db 3061 ctaagagtcgagaatgggtaccgtgccaatccctaataataatagagaagaagagaaactca 3120
Qy 3121 tgaatgggttctcctcaaaagaaaacgaatatattgggacaagaatcgaaaaaattggact 3180
Db 3121 tgaatgggttctcctcaaaagaaaacgaatatattgggacaagaatcgaaaaaattggact 3180
Qy 3181 cattggaaaaaagttaaaattttttgttccaattccaatgttacaacattttcttgaaaaa 3240
Db 3181 cattggaaaaaagttaaaattttttgttccaattccaatgttacaacattttcttgaaaaa 3240
Qy 3241 cgccctaacaagaagaagatggaaatcaatttggtaataagaaataaaagtttgaattatagaa 3300
Db 3241 cgccctaacaagaagaagatggaaatcaatttggtaataagaaataaaagtttgaattatagaa 3300
Qy 3301 aaaaagaacagattattctcttataaaaataaataattc 3337
Db 3301 aaaaagaacagattattctcttataaaaataaataattc 3337

RESULT 3
AAH76457
ID AAH76457 standard; DNA; 3337 BP.
AC AAH76457;
XX
XX
XX 22-OCT-2001 (first entry)
XX Plasmodium vivax ESP-1 DNA.
DE
XX Plasmodium vivax; ESP-1; erythrocyte secreted protein-1; pVESP-1;
KW species-specific; malarial peptide antigen; infection; diagnosis;
KW malaria; ds.
XX
OS Plasmodium vivax.
XX
FH Key Location/Qualifiers
CDS 1..3197
FT /*tag= a
FT /*product= "ESP-1"
FT exon 1..91
FT /*tag= b
FT /*number= 1
FT Intron 92..230
FT /*tag= c
FT /*number= 1
FT exon 231..3197
FT /*tag= d
FT /*number= 2
XX
XX
PN US6231861-B1.
PD 15-MAY-2001.
XX
PF 05-JUN-1998; 98US-0092458.

Db 1621 agcctgtcaaaccaatggtgcgtccacgcgcagatgaaacttattctgtgtatatttag 1680
QY 1681 ataacgatttaacgtatgcagacattacatctctttgagcattatttaacaaatcctca 1740
Db 1681 ataacgatttaacgtatgcagacattacatctctttgagcattatttaacaaatcctca 1740
QY 1741 aggatcctgatgcagagaggtgttaacagttaccatcaaaaggaagcacctgtacaagtac 1800
Db 1741 aggatcctgatgcagagaggtgttaacagttaccatcaaaaggaagcacctgtacaagtac 1800
QY 1801 cagtgagtagggccgcgcgaagaagtcccaacggaagaattgtgcaactccaagaag 1860
Db 1801 cagtgagtagggccgcgcgaagaagtcccaacggaagaattgtgcaactccaagaag 1860
QY 1861 agatttcgaattagaaggaactgcgaagctccagaggaaggaattagttattagaag 1920
Db 1861 agatttcgaattagaaggaactgcgaagctccagaggaaggaattagttattagaag 1920
QY 1921 ggaaggaacacaaacggaagaagccaaagagaagagagccaaacagagaagtgc 1980
Db 1921 ggaaggaacacaaacggaagaagccaaagagaagagagccaaacagagaagtgc 1980
QY 1981 cagaagaacaattagaggaactccagaggaagcagatttcgaattagaagaacacagagag 2040
Db 1981 cagaagaacaattagaggaactccagaggaagcagatttcgaattagaagaacacagagag 2040
QY 2041 aagaagtagaagaacccgtagaggcggaagaactcgaagaagagagaagtgcgaagag 2100
Db 2041 aagaagtagaagaacccgtagaggcggaagaactcgaagaagagagaagtgcgaagag 2100
QY 2101 tacctgcagaagttagaagaagtgcgaaggttacctgcagaagttagaagaagtgcgaagag 2160
Db 2101 tacctgcagaagttagaagaagtgcgaaggttacctgcagaagttagaagaagtgcgaagag 2160
QY 2161 taccagaagaagttagaagaaggttaccgcgaagaagtagaagaagtgcgaaggttaccagaag 2220
Db 2161 taccagaagaagttagaagaaggttaccgcgaagaagtagaagaagtgcgaaggttaccagaag 2220
QY 2221 aegtgaaagaggtaccagaagaagtgcgaaggttaccagaagaagtgcgaaggttaccag 2280
Db 2221 aegtgaaagaggtaccagaagaagtgcgaaggttaccagaagaagtgcgaaggttaccag 2280
QY 2281 aagaagtgcgaagaagtgcgaagaagttagaagaagttagacagcgttagtagaagttag 2340
Db 2281 aagaagtgcgaagaagtgcgaagaagttagaagaagttagacagcgttagtagaagttag 2340
QY 2341 aagttaccagcgttagtagaagaagaaggtccagaagaagttagaagaagaagaagaag 2400
Db 2341 aagttaccagcgttagtagaagaagaaggtccagaagaagttagaagaagaagaagaag 2400
QY 2401 aagaaccagtagagaagaagaagtgtatcaaatagtaataccatcggagaagatatac 2460
Db 2401 aagaaccagtagagaagaagaagtgtatcaaatagtaataccatcggagaagatatac 2460
QY 2461 aattagacaacacaaagaagaacttaggctcgtggaatttttatctatcatcgcacatgc 2520
Db 2461 aattagacaacacaaagaagaacttaggctcgtggaatttttatctatcatcgcacatgc 2520
QY 2521 actaccaagaagcttccaaaggaatttatgcgaagaagaagaagaactgcaggttatccat 2580
Db 2521 actaccaagaagcttccaaaggaatttatgcgaagaagaagaagaactgcaggttatccat 2580
QY 2581 tgaacacagaagaatttgcgaagaagaattcacaactcagaatggtcacaattcattc 2640
Db 2581 tgaacacagaagaatttgcgaagaagaattcacaactcagaatggtcacaattcattc 2640
QY 2641 aagggcctagaagcgcactgggaacgattagaagtgcgttaataaaggcttagagaagaat 2700
Db 2641 aagggcctagaagcgcactgggaacgattagaagtgcgttaataaaggcttagagaagaat 2700
QY 2701 ggaatggaacaaagaataaagaatggcgtgcgtgcgttcgcttaattgaaataaatggt 2760
Db 2701 ggaatggaacaaagaataaagaatggcgtgcgtgcgttcgcttaattgaaataaatggt 2760

QY 2761 cagaatatagtcaaatttcaacaaaggaagaccagctggttttgagaaacagagct 2820
Db 2761 cagaatatagtcaaatttcaacaaaggaagaccagctggttttgagaaacagagct 2820
QY 2821 ggagcgacgagaaatgcgaagaatggttttaaagcagaagtcaaatcccaaatgattcac 2880
Db 2821 ggagcgacgagaaatgcgaagaatggttttaaagcagaagtcaaatcccaaatgattcac 2880
QY 2881 acttgaacaaatggatgaacgacactcattccaatttattaaaattctctggaagata 2940
Db 2881 acttgaacaaatggatgaacgacactcattccaatttattaaaattctctggaagata 2940
QY 2941 tctcacaatttgaaacaaagaaacccaaagaaatgggtaataatcactctggaacaaagc 3000
Db 2941 tctcacaatttgaaacaaagaaacccaaagaaatgggtaataatcactctggaacaaagc 3000
QY 3001 aacgggggttatgggttcgaatcatttggaagttagaccacatcaaaattattaaatgtgg 3060
Db 3001 aacgggggttatgggttcgaatcatttggaagttagaccacatcaaaattattaaatgtgg 3060
QY 3061 ctgaagagtcgaagaatggtaccgtgccaatcctcaataataatagagaagaagaactca 3120
Db 3061 ctgaagagtcgaagaatggtaccgtgccaatcctcaataataatagagaagaagaactca 3120
QY 3121 tgaaaatgggttctcctaaaaagaaacgaataatttaggacaaagaaatggaaaaatggact 3180
Db 3121 tgaaaatgggttctcctaaaaagaaacgaataatttaggacaaagaaatggaaaaatggact 3180
QY 3181 cattggaaaaaaagttaaattttttgggttcgaattcgaatgcataacatttcttgaaaaa 3240
Db 3181 cattggaaaaaaagttaaattttttgggttcgaattcgaatgcataacatttcttgaaaaa 3240
QY 3241 cgcctaaacaaagaagaatggaatcgaatgttgaataaataaaagttagaattatagaa 3300
Db 3241 cgcctaaacaaagaagaatggaatcgaatgttgaataaataaaagttagaattatagaa 3300
QY 3301 aaaaagacagattattctcttataaaataaaataattc 3337
Db 3301 aaaaagacagattattctcttataaaataaaataattc 3337

RESULT 4
AAT05868
ID AAT05868 standard; DNA; 3399 BP.
XX
AC AAT05868;
XX
DT 14-AUG-1996 (first entry)
XX
DE Chicken leucocytozoan DNA encoding immunogenic protein for vaccines.
KW Chicken leucocytozoan; immunogen; recombinant vaccine; protection;
immunisation; vaccination; ss.
XX
OS Chicken leucocytozoan.
FH Key Location/Qualifiers
FT CDS 1..3399
FT /*tag= a
FT misc_feature 1150..3218
FT /*tag= b
FT /note= "fragment referred to in the claims, for
FT use as insert in a recombinant vaccine
FT against chicken leucocytozoan disease"
XX
PN JP07284392-A.
XX
PD 31-OCT-1995.
XX
PF 19-APR-1994; 94JP-0080643.
XX 19-APR-1994; 94JP-0080643.

Db 803 ggaagcagaagagggatgcagaagaagaagaagggaagtcggaagaagaaggaggg 862
Qy 1945 agccaagagagagagcacaagaagga-gaagtcccagaagaagaattagagcaact 2003
Db 863 ggaagaggaacaaatatgaaggggggggggggggggggggggggggggggggggg 922
Qy 2004 ccagagcagatttcgaattagaagaaccaacagaggaagaagtagaagaacccgtagag 2063
Db 923 gaaagagagaggaagaagaagagcagaagcaggggaaggggagaggaagaaggagg 982
Qy 2064 ggcgaagaactcagaagaagaggaagtcggaagaggtacctgcagaagtagaagaagtg 2123
Db 983 gaagcagggcgaggaagaagggggggggaagaagggggaagaggtaggggggaagag 1042
Qy 2124 gaagaggtacctcagaagtagaagaagtcggaagaggtaccagaagaagtagaagaggt 2183
Db 1043 ggcgaaggagaggaagaagaagggggggggggaagaaggagagggaggggggaagag 1102
Qy 2184 ccgcgaagtagaagaagtcggaagaggtaccagaagaagtcggaaggtaccagaagaa 2243
Db 1103 gaaggaaggaggggaggaaggaaggagaggaagggaaggtgagagggaaggaagta 1162
Qy 2244 gtggaagaggtaccagaagaagtcggaagaggtaccagaagaagtcggaagaagtggaa 2303
Db 1163 ggggagagaggaagaacagaagggaagggggacaaaggaagagagagggagggagaa 1222
Qy 2304 gtagaagaagtagaggtaccagcggtagtagaagtagaagtagaccagcggtagtaga 2363
Db 1223 gagggagaggaagcagagtcggaagtcggaagtagagggaggaagagacactgaagag 1282
Qy 2364 gaggtgccagaagaagtagaagaagaagaagaaggaaggaagcagtagaggaagag 2421
Db 1283 ggaataatgagagacagagagtagagagagtgaaagagtgaaacacagggaagcgggtg 1340

RESULT 12
AAD07132
ID AAD07132 standard; cDNA; 2803 BP.
AC AAD07132;
XX AAD07132;
DT 06-AUG-2001 (first entry)
DE Canine retinitis pigmentosa GTPase regulator (RPGR) mutant #2 cDNA.
KW Dog; X-linked progressive retinal atrophy 2; XLPRA2; genetic marker;
KW retinitis pigmentosa GTPase regulator; RPGR; Siberian Husky; Samoyed;
KW Miniature Schnauzer; mutant; mutuin; ss.
XX
OS Canis familiaris.
OS Synthetic.
FH Key
FT CDS
FT 1..1149
FT /tag- a
FT /product- "Canine retinitis pigmentosa GTPase
FT regulator mutant"
FT /note- "CDS does not include start codon"
FT /partial
FT mutation
FT replace (931..932, AGAG)
FT /tag- b
FT /note- "This deletion results in the change of
FT amino acids and ends in a premature stop codon"
XX
PN WO200138578-A1.
XX
XX 31-MAY-2001.
XX
XX 21-NOV-2000; 2000WO-US31940.
XX
XX 24-NOV-1999; 99US-0167365.
XX

PA (CORR) CORNELL RES FOUND INC.
XX Aguirre GD, Acland GM, Zhang Q, Ray K, Zeiss CJ;
PI WPI: 2001-367707/38.
DR P-PSDB; AAE02399.
XX
PT Identifying dogs with or carrying X-linked progressive retinal atrophy
PT by detecting retinitis pigmentosa GTPase regulator gene mutation,
PT useful when breeding Husky, Samoyed and Miniature Schnauzer -
XX
PS Claim 75; Page 34-35; 88pp; English.
XX
CC The invention relates to a method for identifying dogs which are
CC genetically normal, are carriers of, or are affected with X-linked
CC progressive retinal atrophy (XLPRA), by testing a biological sample with
CC genetic markers that co-segregate with a XLPRA gene locus. The invention
CC also relates to canine retinitis pigmentosa GTPase regulator (RPGR), RPGR
CC mutants and their corresponding nucleic acid molecules. The mutated RPGR
CC genes are responsible for the XLPRA in dogs. Methods are used to select
CC dogs for breeding so that dogs carrying the mutated locus are eliminated
CC from the breeding stock. The method particularly applies to Siberian
CC Husky, Samoyed and Miniature Schnauzer breeds, or any other breed where
CC the disease is X-linked. XLPRA1 type is identified in Siberian Huskies,
CC and Samoyeds, while XLPRA2 type is identified in Miniature Schnauzers.
CC The present cDNA sequence is the mutant open reading frame (ORF) 15
CC encoding Canine retinitis pigmentosa GTPase regulator (RPGR) mutant
CC found in XLPRA2-affected dogs. This mutant is obtained by deleting
CC 'GA' nucleotide bases from position 932 to 933 of the wild-type
CC canine RPGR cDNA.
XX
SQ Sequence 2803 BP; 1028 A; 287 C; 939 G; 549 T; 0 other;
Query Match 6.3%; Score 211; DB 22; Length 2803;
Best Local Similarity 48.3%; Pred. No. 8; 9e-26;
Matches 622; Conservative 0; Mismatches 665; Indels 2; Gaps 1;
Qy 1225 gaaggagaacacagaaacgtagaggagagaagaacagagctgcagaaggagaag 1284
Db 143 gagagaaaaacctagagaggagagggcccgaggaacaaaggaggggaacaggccatc 202
Qy 1285 tatcagaagaactccagaaggagagaagaagtagtagggccaactccagagagtagttct 1344
Db 203 ggaagaagaacaaacaggaagagagagaggggaagggagcaagagatgagaag 262
Qy 1345 cattagatggaactacattagaagaacacgaacacgaagagaggaagaacacgtag 1404
Db 263 aaggggagggagagagggggaagagaccgggaagagggagagggaggaaggca 322
Qy 1405 agggagaagaacccgtagagggagaagaacccgtagagggagaagagctgcagaaggag 1464
Db 323 aggaggaaggagaatgaaagagaaggacagggagagggggaagaagaaggagagg 382
Qy 1465 aagaagagttagaggcaactccagaggatgacttccaattagaaagacacacagagag 1524
Db 383 aggaagcagaggagaagaagaagaagcagcagaggggggagggagggagggggaag 442
Qy 1525 gagaaggggaagagaagaaggaaggaaggaaggaagagagcgttagtagcagtgccag 1584
Db 443 aggaggaaggagagaggggggaaggaagaagaagggggggaagggggaaggagag 502
Qy 1585 tagtgccgaacccgttagaagtagtgactctctcagcctgcacacacacacaggtgctc 1644
Db 503 cagagaaagaagaagcagaagaagtagaggaaggaagggggggaattggaagagg 562
Qy 1645 caacggcagatgaactttatttcgttgatatcttagataacgatttaacgtatgcagaca 1704
Db 563 aaggggggggggaattagaggaagagaggggaaagggggaagaagagagggggaattgg 622
Qy 1705 ttacatcccttgagccattatttaacacaaatcctcaaggtatctgatcgaggagagctg 1764
Db 623 agggggaagaggagggggggaagaggaaggggaggaaggggaggggaggggaggg 682

xx The present invention relates to a method for diagnosing disease or
cc predisposition to a disease, associated with a disease causing
cc mutations in a retinitis pigmentosa GTPase regulator (RPGR) gene
cc involves genotyping a RPGR gene, and determining whether the genotype
cc comprises a disease causing mutation, where the risk genotype is
cc present within open reading frame (ORF)15 of the RPGR gene. The method
cc is useful for detecting a certain disease state e.g., X-linked
cc retinitis pigmentosa (XLRP). The kit is useful for detecting and
cc measuring disease causing mutations in biological fluids and tissues
cc and for localising mutation in tissues. The mutant RPGR gene is useful
cc in gene therapy techniques and for screening agents capable of
cc affecting the expression of the sequences and/or the biological
cc activity of mutant RPGR. They are preferably useful for identifying
cc agonists and antagonists of RPGR. The mutant RPGR gene is also useful
cc in identification of potential pharmaceutical targets in the high
cc throughput screening assays and forensic analysis. The present sequence
cc is the mutational hot spot sequence of human RPGR exon ORF15 DNA.
xx
SQ Sequence 1050 BP; 428 A; 15 C; 595 G; 12 T; 0 other;

Query Match 6.0%; Score 200.6; DB 22; Length 1050;
Best Local Similarity 51.1%; Pred. No. 3.9e-24;
Matches 525; Conservative 0; Mismatches 494; Indels 8; Gaps 2;

QY 1406 gggagaagaaccgtagagggagaagaaccgtagaggagaagaagctgcagaaggaga 1465
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
1 gggagcaggccatcagaagaaggaaaccacaagatggaggag ----ggaggggaggga 55
QY 1466 agaaagtttagggccaactccagaggtgacttccaattagaagaaccatcaggagaagg 1525
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
56 ggaagcatggaggaagaagaggagggagacagagagaaggaaggaagaaggagg 115
QY 1526 aqaagggaagagagaagagaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 1585
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
116 aqaagggaagaggaagggaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 175
QY 1586 agtgccgaacccgttagagtagtgactctctcagccgtcaaaccaatgctgcctcc 1645
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
176 agagaggaaggaaggaaggaagggaggaaggaaggaaggaaggaaggaaggaacca 235
QY 1646 aacggcagatgaactttatcttgatctttagataacagatttaacgtatgcagacat 1705
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
236 aggaagggggaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 295
QY 1706 tacatctcttgagccattttaacaaatcctcaagatcctgatgcagagagagcgtgt 1765
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
296 agtagagggagggagtagagggggaaggaaggaaggaaggaaggaaggaaggaaggg 355
QY 1766 aacagtaccatcaagaggacacctgtacaagtaccagtggcagtagggccgcgcaga 1825
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
356 tgagggggagagaggaagggagggggaagagaggaaggaagggggggaagagaggga 415
QY 1826 agtgcacaggaagaattgatgcactccaagagacgatctcgaattagaaggaaactgc 1885
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
416 aggaagaagggaagggggaaggaaggaaggaaggaaggaaggaaggaaggaagggaggga 475
QY 1886 aqaagctccagaggaaggaattagtattagaaggaaggaaggaaggaaggaaggaagga 1945
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
476 aggaagaagggggaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggg 535
QY 1946 gccagaagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2005
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
536 ggaagggaaggaaggaaggaagggagggaggaaggaaggaaggaaggaagggaggaga 595
QY 2006 agaggaagatttcgaattagaagaaccacaggaaggaagaagtagagaacccgtgaggg 2065
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
596 agaggaaggaaggaaggaagggaggaaggaaggaaggaaggaaggaaggaagggaggga 655
QY 2066 cgaagaactgcagaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaagga 2125
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
656 agaggaaggaaggaaggaaggggagggatggagaagggggagggggaaggaaggaaggg 715

QY 2126 agaggtacctgcagaagtagaagaagtgaaaggtaccagaagaagtagaagaaggtacc 2185
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
716 agaattggggggggaagaggaaggaagggggggggggggggggggggggaagggga 775
QY 2186 ccagaagtagaagaagtgaaaggtaccagaagaagtggaagaggtaccagaagaaggt 2245
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
776 agggggaggaaggaagggggggggggggggggggggggggggggggggggaagga 835
QY 2246 ggaagaggtaccagaagaagtgaaaggtaccagaagaagtggaagaggtggaagaaggt 2305
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
836 ggaagggg---aagaagaagggggggaaggaaggaaggaaggaaggaagggaggaga 892
QY 2306 aqaagaagtagaggtaccagcggtagtagaagtagaagtagaccagcggtagtagaagaaga 2365
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
893 aggggaggaagaaggaaggggggggggggggggggggggggggggggggggaggagg 952
QY 2366 ggtgccagaagaagtagaagaagaagaagaaggaaggaaggaaggaaggaaggaaggtgt 2425
Db | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
953 ggaagagaggaaggaagggaggaggaggaggaggaggaggaggaggaggaggaggagg 1012
QY 2426 attacaa 2432
Db | | | |
1013 agaagaa 1019
RESULT 14
AAS90688
ID AAS90688 standard; cDNA; 621 BP.
XX
AC AAS90688;
XX
DT 13-FEB-2002 (first entry)
XX
DE DNA encoding novel human diagnostic protein #26492.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR P-PSDB; ABG26501.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 1; SEQ ID No 26492; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as

[illegible][illegible]

CDS 8699..11236
 /*tag= b
 /product= glycoprotein B
 complement (17261..17875)
 /*tag= c
 /product= interleukin 6
 complement (21548..21832)
 /*tag= d
 /product= macrophage inflammatory protein II
 complement (27137..27424)
 /*tag= e
 /product= interferon regulatory factor 1
 28661..29741
 /*tag= f
 /product= protein T1.1
 complement (58976..60175)
 /*tag= g
 /product= glycoprotein M
 complement (69412..69915)
 /*tag= h
 /product= glycoprotein L
 complement (88410..88910)
 /*tag= i
 /product= interferon regulatory factor 2
 89600..90541
 /*tag= j
 /product= interferon regulatory factor 3
 90173..90643
 /*tag= k
 /product= glycoprotein X
 complement (93636..94127)
 /*tag= l
 /product= interferon regulatory factor 4
 complement (111931..112443)
 /*tag= m
 /product= capsid protein IV
 complement (123808..127296)
 /*tag= n
 /product= immediate early protein
 WO9804576-A1.
 05-FEB-1998.
 22-JUL-1997: 97WO-US13346.
 29-NOV-1996: 96US-0757669.
 25-JUL-1996: 96US-0686243.
 25-JUL-1996: 96US-0686349.
 25-JUL-1996: 96US-0686350.
 25-JUL-1996: 96US-0687253.
 25-JUL-1996: 96US-0688814.
 05-SEP-1996: 96US-0708678.
 10-OCT-1996: 96US-0728323.
 13-NOV-1996: 96US-0747887.
 13-NOV-1996: 96US-0748640.
 (UYCO) UNIV COLUMBIA NEW YORK.
 Bohenzky RA, Chang Y, Edelman IS, Moore PS, Russo JJ;
 WPI; 1998-130615/12.
 New nucleic acid encoding Kaposi's sarcoma associated herpes virus
 proteins - useful for, e.g. detecting levels of HHV8 in, and
 preparation of vaccines for treatment of, HIV patients
 Example 2; Page 135-203; 230pp; English.
 This sequence represents the long unique region and terminal repeat of
 the Kaposi's sarcoma-associated herpes virus (KSHV). KSHV is also known
 as human herpes virus 8 (HHV8). This sequence contains the DNAs of the
 invention which encode KSHV polypeptides selected from: (a) viral

CC macrophage inflammatory protein (MIP) II; (b) viral interleukin-6 (IL-6);
 CC (c) viral Irf 1; (d) complement-binding protein; glycoproteins B, M or L;
 CC (d) capsid protein IV encoded by ORF65; and (e) immediate early protein
 encoded by ORF73. Labelled probes for the nucleic acid, proteins encoded
 by it, and antibodies (Ab) specific for the proteins are useful for
 CC detecting HHV8, specifically for diagnosis of Kaposi's sarcoma, in body
 fluids or tissue samples. HHV8 infections can be treated with antisense
 CC or triplex forming molecules or agents that bind specifically to the
 CC protein. Ab may be used for prophylaxis or treatment of HHV8 infection,
 CC while the protein can be used in protective vaccines. Ab may also be used
 CC to differentiate between lymphomas, and HHV8 may be implicated in many
 CC other lymphoproliferative diseases such as lymphomas, leukaemia,
 CC splenomegaly and mycosis fungoides. Cells and animals containing the
 CC nucleic acid are useful for drug screening. HHV8-derived peptides can be
 CC used as targets for antiviral drugs, e.g. dihydrofolate reductase gene
 CC can be inhibited with methotrexate. These can also be used to determine
 CC the immune status of a patient infected with HIV. HHV8 derived protein
 CC viral MIP III may be used as an anti-inflammatory agent for,
 CC e.g. treating rheumatoid arthritis. This sequence is stated as containing
 CC 81 open reading frames.
 XX
 SQ Sequence 137507 BP; 32579 A; 37795 C; 35758 G; 31375 T; 0 other;
 Query Match 6.0%; Score 199; DB 19; Length 137507;
 Best Local Similarity 46.9%; Pred. No. 1.6e-23;
 Matches 685; Conservative 0; Mismatches 770; Indels 4; Gaps 2;
 QY 1011 gaagagccgcgaagtaccagcggaagaattagatgccatccacagaggatgacttcgca 1070
 DB 126270 GAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 126211
 QY 1071 ttgatgtttacagaatctcccgaggaagtagaattagattagatgaagagcaactgaa 1130
 DB 126210 GAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAGGAGGATGACGAG 126152
 QY 1131 gaagaatcaacggaagtgggaccaaagggaagggaccaaagggaaggaatagatgccact 1190
 DB 126151 GGAGGATGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 126092
 QY 1191 ccagagatggtatttcgattagacgaactgcagagaggaagaaacgaacgtatag 1250
 DB 126091 CGAGGAGGATGACGATGATGAGGACCAATGAGGACGAGGATGACGAGGAGGAGGAGGAG 126032
 QY 1251 ggagaagaaacagaagaagctgcagaaggagaagtagatcagaagaacacccaggaaggagaa 1310
 DB 126031 GAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 125972
 QY 1311 gaagagtttagaggaactccagaggatgatttcgattagatgaactagatagaagaa 1370
 DB 125971 ACAGCAGCAGGAGGCCAACACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGGAGGCC 125912
 QY 1371 accgaagaaactcagaagagagaagaacccgtatagaggag---aagaaacgtatagaggga 1427
 DB 125911 CTTGCAGGAGGCCACACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGGAGGCCACAG 125852
 QY 1428 gaagaaacccgtagaggagagaagaagctgcagaaggagaagaagagtagtagggcaactcca 1487
 DB 125851 GCAGGAGGCCAACACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGGAGGCCACACAGCA 125792
 QY 1488 gaggatgaacttccaattagaagaaccatcaggagaagagaagggggaagagagagaa 1547
 DB 125791 GGAGGCCAACACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGCA 125732
 QY 1548 ggagaaggaagagaagaaagcgttagtagcagtagtccagtagtgcacaaacggtagaagta 1607
 DB 125731 GGAGGCCACACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGCA 125672
 QY 1608 gtgactcctctcagccctgtcaaaccaatggtcgtcccaacgagatgaactttattc 1667
 DB 125671 GGAGGCCACACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGGAGGCCACAGCAGCA 125612
 QY 1668 gttgatattctagataacgatttaacgtatgcagacaattacattctcttgagccatttt 1727

Db 2596 CAGCGGAAGATCCT-----TCAGCTGGGAACACAGAGCGGCAGATGATCCATCGGCTG 2543
QY 1039 aattagatgcactccagagatgaacttcgcatattagattacagaatctcccgaggaag 1098
Db 2542 GAGCACCAGAGCGGAGTTGTGACATCAGCAGGAGAGCCAGGAGCGGGAAGATCCGT 2483
QY 1099 tagaattagattagatgaagaggaactgaagaatacaacggaagtgggaccaacgg 1158
Db 2482 CAGCATCAGCAGGAGAGCCAGGAGAGCGGGAAGATCCGTGAGTGGAGCGCCAGGAGCGG 2423
QY 1159 aagaagaccacacgaagaattagatgccactccaagatgatcttcattagacga 1218
Db 2422 CAGATGATCCATCGCTGGAGACACAGGAGCGGAGTTG-TGACATCAGCAGGAGAGCCCA 2364
QY 1219 actgcagaaggaagaacagagaacacgttagagggagaagaacacagaagaagctgcagaag 1278
Db 2363 GGAGCAGCGGAAGATCCGTGAG--CTGGAGCGCCAGGAGCGGAGATGATCCATCGGCTG 2306
QY 1279 gagaagtatcagaagaactccgaaggaagaagaagattagagggaactccagagatg 1338
Db 2305 GAGCACCAGGAGCGGAGTTGTGACATCAGCAGGAGCGGAGGAGCGGGAAGATCCCTT 2246
QY 1339 atttcgcatattagatgaactacattagaagaacccagaagaactgcagaaggagaagaa 1398
Db 2245 CAGCTGAGAAACAGGAGCGGAGATGATCCATCGCTGGAGCACCAGGAGCGGAGTTG 2186
QY 1399 ccgtagagggaagaacccgttagagggagaagaacccgttagaggagaagaagctgcag 1458
Db 2185 TGACATCAGCAGGAGAGCCAGGAGCAGCGGAGAGATCGTCACTGGAGCGCAGGAGCGG 2126
QY 1459 aaggagaagaagtttagaggaactccagagagatgaactccaattagaagaaccatcag 1518
Db 2125 GAGAAGATCCATCAGCTGGAGCGCCAGGAGCGGAGTTGTGACATCAGCAGGAGAGCCAG 2066
QY 1519 gagaagagagaaggagaaggaagggaaggaaggaaggaagcgttagtagcag 1578
Db 2065 GAGCAGCGGAAGATCCGTCACTGAGTGGAGCGCCAGGAGCGGAGTTGTGACATCAGCAGGAG 2006
QY 1579 tgcagtagtgccgaaccggtagaagtagtgaactcctgcagcctgtcaaacccaatgg 1638
Db 2005 AGCCAGGAGCAGCGGAAGATCCGTCACTGAGTGGAGCGCCAGGAGCGGAGATGATCCATCGG 1946
QY 1639 tcgctcaacgagcagatgaactttattcgttgatattcttagataacgatttaacgtatg 1998
Db 1945 CTGGAGCACCAGGAG-----CGGCAGTTGTGACATCAGCAGGAGCGCCAGGAGCAGCGG 1892
QY 1699 cagacattacatctttgagccattatttaacaaatccctcaaggatcctcgtatgcaggag 1758
Db 1891 AAGATCCTTCACTGAGTGGAGAACAGGAGCGGAGATGATCCATCGGCTGGAGCACCAGGAG 1832
QY 1759 aggc---tgtaacagtaccatcaaaaggaagcacctgtacaagtaccagtggcagtagggc 1815
Db 1831 CGGCAGTTGTGACATCAGCAGGAGCGCCAGGAGCGGAGATGATCCATCAGCTGGAGCGC 1772
QY 1816 ccgcgaagaagtgcgaacggaagaatttgatgaactccaagaggagcatttcgaattag 1875
Db 1771 CAGGAGCGGAGTTGTGACATCAGCAGGAGAGCAGGAGCGGAGATGATCCATCGCTGAGCTG 1712
QY 1876 aaggaactgcagaagctccagaggaaggaatttagattagaaggagaaggaagaa 1935
Db 1711 GAGCGCAGGAGCGGAGATGATCCATCAGTGGAGCGCCAGGAGCGGAGTTGTGACAT 1652
QY 1936 cgaagaagaagcgaagaagagagccaaacagaagaaggaagtgccagaagaagaattag 1995
Db 1651 CAGCAGGAGAGCCAGGAGCAGCGGAAGATCCGTGAGTGGAGCGCCAGGAGCGGAGATG 1592
QY 1996 agccaactccagaggaagcatttcgaattagaagaacccaacagagaagaagtagagaaaa 2055
Db 1591 ATCCATCGGCTGGAGCACCAGGAGCGGAGTTGTGACATCAGCAGGAGCGCCAGGAGCAG 1532
QY 2056 ccgttagagggcgaagaactgcagaaggaagaagtggaagaggtacctgcagaagtag 2115
Db 1531 CGGAAGATCCTTTCAG---CTGGAGAACCCAGGAGCGGAGATGATCCATCGGCTGGAGCAC 1475

QY 2116 aagaagtgaagagggtacctgcagaagtgaagaagtgaagagggtaccagaagaagtag 2175
Db 1474 CAGGAGCGGCAGTTGTGACATCAGCAGGAGAGCCAGGAGCGGGAAGATCCGTGAGCTG 1415
QY 2176 aagaggtaccgccagagaagtgaagaagtgaagagggtaccagaagaagtgaagaaggtac 2235
Db 1414 GAGCGCCAGAGCGGAGATGATCCATCGCTGGAGCACCAGGAGCGGAGTTGTGACAT 1355
QY 2236 cagaagaagtgggaaggttaccagaagaagtgggaaggttaccagaagaagtgggaagaag 2295
Db 1354 CAGCAGGAGCGCCAGCAGGAGCGGAAGATCCCTTCAGCTGGAGAACCCAGGAGCGGAGATG 1295
QY 2296 tgaagaagttagaagaagttagaggttaccagcggttagaagttagaagtaccagcggttag 2355
Db 1294 ATCCATCGGCTGGAGCACCAGGAGCGGAGTTGTGACATCAGCAGGAGAGCCAGGAGCAG 1235
QY 2356 tagaagaagaaggtgcagagaagaagtgaagaagaagaagaagaaggaaggaaccagtagagg 2415
Db 1234 CGGAAGATCCGTGAGTGGAGCGCCAGGAGCGGAGATGATCCATCGGCTGGAGCACCAG 1175
QY 2416 aagaagatgtattacaaattagtaataaccatcggaagaag 2454
Db 1174 GAGCGGCAGTTGTGACATCAGCAGGAGAGCCAGGAGCAG 1136

RESULT 20

ABL05186/C
ID ABL05186 standard; cDNA; 9516 BP.

XX ABL05186;

XX XX 26-MAR-2002 (first entry)

DT Drosophila melanogaster expressed polynucleotide SEQ ID NO 10040.

DE Drosophila; developmental biology; cell signalling; insecticide;

XX pharmaceutical; gene; ss.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR P-PSDB; ABB61083.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

PS Claim 1; SEQ ID NO 10040; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO


```
RESULT 21
AAQ03875
ID AAQ03875 standard; DNA; 3095 BP.
XX AC AAQ03875;
XX DT 24-AUG-1990 (first entry)
XX DE Sequence encoding carboxylic terminal part of native GLURP.
XX KW Plasmodium falciparum; antigen; malaria; vaccine; GLURP;ss
XX OS Plasmodium falciparum.
XX FH Key Location/Qualifiers
XX CD 1..2352
XX FT /tag= a
XX FT /product=GLURP
XX PN W09022811-A.
XX PD 22-MAR-1990.
XX PF 18-SEP-1989; 89WO-0000218.
XX PR 03-MAR-1989; 89US-0218885.
XX PR 03-MAR-1989; 89DK-0005191.
XX PA (STAT-) STATENS SERUMINST.
XX PI Driegiel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A;
XX PI Jakobsen PH;
XX DR WPI: 1990-115998/15.
XX DR P-PSDB; AAR05804.
XX PT Polypeptide(s) derived from Plasmodium falciparum antigen - used in
XX PT vaccines and in production of antibodies, for diagnosis and
XX PT therapy of malaria.
XX PS Disclosure; Fig 7; 108pp; English.
XX CC An open reading frame of 2349 bps extends from the 5' terminal end of the
XX CC insert to a "TAA" stop codon. It is longest ORF found in the sequence.
XX CC Sequence displays some of the characteristics of other malaria nucleic
XX CC acid sequences : tandemly repeated motifs, high AT content and a
XX CC corresponding preference for codons containing these bases, and a high
XX CC content of codons for glutamate. Three major repetitive sequences are:
XX CC one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another
XX CC motif from bp 477 to bp 521 is repeated tandemly twice from bp 522 to bp
XX CC 566 and from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is
XX CC repeated tandemly 11 times. This last repetitive region consists of 360bp
XX CC repeats differing only in 3 bases GAT coding for aspartate. This region
XX CC is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content
XX CC of the coding part of the insert is on average 30%, and of the non-coding
XX CC 3' terminal 11%.
XX SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T; 0 other;

Query Match 5.1%; Score 171.8; DB 11; Length 3095;
Best Local Similarity 44.8%; Pred. No. 2.4e-19;
Matches 846; Conservative 0; Mismatches 1022; Indels 21; Gaps 4;

QY 597 gaaactggcatttgaagcagggggaagaactggcgatttggagcaggggaagaagct 656
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
DB 16 gaaaaagcagcatgaagcagctgaaatgaagaagtgtcttgaagaagccatcat 75
QY 657 ggtgatttgaagcagggggaagaactggcgatttggagcaggggggaagaactggagat 716
    || || || || || || || || || || || || || || || || || || || ||
DB 76 gaagaaattgtacctgacaaataatgaagaatcaggtgaagtaaatagttgataat 135
QY 717 gcggaaactgaagaagcagaactggagatgcggaactggaataatggagcaactgtgat 776
```

```
DB 136 gatgaaggtgttttgaagaagctcatcatgaaaaattttcatctgaagtaactct 195
QY 777 gtagacacagaagatgattcagctgatggagcagaaaaaacatcatcttctctcaagaa 836
    || || || || || || || || || || || || || || || || || || || ||
DB 196 gaattaaatgaataatgaatttgtgaatctgacaaaatgttaactgaacctgtgaacat 255
QY 837 aatgtacaacctccgatagtaaatgatgcctcttttggaaagtattttggataaagata 896
    || || || || || || || || || || || || || || || || || || || ||
DB 256 gaagaa-----gtgtatctgaagaagcaacctgaaccagctgaaaatgaagaagt 309
QY 897 attttgatcatattaaagatttcgagccactatttcgacaaatgttg9gggtactgct 956
    || || || || || || || || || || || || || || || || || || || ||
DB 310 agtatagaagaagctcatcagggaataattgtacctgacacaaatgatgaagaatcaggt 369
QY 957 aaacatgtttacgggacaagaattgccaatgaaacctgtaccattaccagtggcagaagag 1016
    || || || || || || || || || || || || || || || || || || || ||
DB 370 gaaagtggattagatgaatgaagaagtgatttggagaacctaatcatgaagaattt 429
QY 1017 ccgcgcgaagtaccagcgggaagaattagatgccactccagaggtgacttcgcattagat 1076
    || || || || || || || || || || || || || || || || || || || ||
DB 430 gaacctgatcaaaatgactctgaatttaagtgaataatgaattagtcagaaaaaagt 489
QY 1077 gttacggaatctcccagagggaagttagaattagttattagatgaagagcgaactgaagaagaa 1136
    || || || || || || || || || || || || || || || || || || || ||
DB 490 gtatctgaaccagctgaacatgtagaattgtatcagaaaaaagtgtatctgaaccagct 549
QY 1137 tcaacggaagtggaccaacgcgaagagaccacgaagaattagatgccactccagag 1196
    || || || || || || || || || || || || || || || || || || || ||
DB 550 gaacacgtgaaattgtatctgaaaaaagtacatccgaa-----ccagctgaaca 599
QY 1197 gatggatttcgcattagacgaaactgcagaagagagaacagaagaacgttagagggagaa 1256
    || || || || || || || || || || || || || || || || || || || ||
DB 600 tgtagaagtgtatctgaacaaagttaataacgaacctccga--aaagaaagtggacca 657
QY 1257 gaaacgagaagaactgcagaagagagaagtatcagaagaacactccagaagagagaagag 1316
    || || || || || || || || || || || || || || || || || || || ||
DB 658 gttccttcaaaacatttgaagaattgaaaagtggatttcaacctaaattgtgac 717
QY 1317 ttagaggccaactccagaggatgttccgatttagatggaaactcattagagaacccgaa 1376
    || || || || || || || || || || || || || || || || || || || ||
DB 718 ctcaataattgaacctaatittgtgactcaacaaatccacaagaacagattgaa 777
QY 1377 gaaactgcagaaggaagaacacgttagagggagaagaacaccttagagggagaagaacc 1436
    || || || || || || || || || || || || || || || || || || || ||
DB 778 ccatctattgtcaaaattgaaaagtctctctgagaagaaataaacatgcaagtgtgat 837
QY 1437 gtgaggggaaagaactgcagaaggaagaagatttagaggcaactccagaggatgac 1496
    || || || || || || || || || || || || || || || || || || || ||
DB 838 cctgaagtataaagaaaaaagaaatgtatctgaagtgttgaagaaaaacaaattccaa 897
QY 1497 ttccaatagaagaacacatcaggagaagggaggaaggggaaggaagggggaagga 1556
    || || || || || || || || || || || || || || || || || || || ||
DB 898 gaatcagtgaagaataatccagtaaatgagatgaatttgaagatgttcacactgaacaa 957
QY 1557 gaaggagaagcgttagtagcagtgccagtagtgccgaaacccggtagaagtagtgaactct 1616
    || || || || || || || || || || || || || || || || || || || ||
DB 958 ttagatttagatcataaaacagttgtaccagaaatagtagaagttagaagaattcctca 1017
QY 1617 gctcagcctgtcaaaccaatgttcgctcccaacgcgcagatgaaactttatctgtgatc 1676
    || || || || || || || || || || || || || || || || || || || ||
DB 1018 gaactacatgaataatgaagtggctcatccagaaatgttgaatttagagggaagtttccct 1077
QY 1677 ttgataaacgatttaacgtatgcagacattacatctctttgagcattattttaacaacatc 1736
    || || || || || || || || || || || || || || || || || || || ||
DB 1078 gaaccaatacaataacgaatttcaagaaataatgaagatgataaaaagtcacatatt 1137
QY 1737 ctcaagatcctgatgcaggagagaggtgttaacagtaccatcaaaagggaagcacctgtacaa 1796
    || || || || || || || || || || || || || || || || || || || ||
DB 1138 cagcatgaataatagaaagttagaagaataactccagaagatgataaaaatgaaaaagtt 1197
QY 1797 gtaccagtggcagtagggcccgccgcaagaagtgccaacgcgaagaattgtgcaactccaa 1856
    || || || || || || || || || || || || || || || || || || || ||
```

Db 1198 gaacatgaaatagtagaagttgaagaaattctagcagaagaataaaatgaaaaggtcaa 1257
 QY 1857 gagagcatttcgaattagaaagaaactgcagaagctccagaggaaggaagaaattagatta 1916
 Db 1258 catgaaatagtagaggtgaagaattctaccagaagatgataaaatgaaaaggtcaa 1317
 QY 1917 gaagggaagagagaacaaacggaagaagagcgaagaaggaaggaagcgaacagaagagaa 1976
 Db 1318 catgaaatagtagaagttgaagaattctaccagaagataaaaaatgaaaaggtcaacat 1377
 QY 1977 gtccagaagaagaattagaggcaactccagagagcagatttcgaattagaagaaccaaca 2036
 Db 1378 gaattgtagaggttagaagaattctaccagaagataaaaaatgaaaaggtgacatgaa 1437
 QY 2037 ggagaagaagtagaagaacccgtagagggcggaagaactgcagaag--agaagaagt 2093
 Db 1438 atagtagaagttgaagaattctaccagaagataaaaaatgaaaaggtcaacatgaata 1497
 QY 2094 gaagaggtacctcagaagtagaagaagtggaagaggtacctcagaagtagaagaagt 2153
 Db 1498 gtgaggttgaagaattctaccagaagataaaaaatgaaaaggttcaacatgaataagta 1557
 QY 2154 gaagaggtaccagaagaagtagaaggttaccgcgagaagtagaagaagtggaagaggt 2213
 Db 1558 gaagttgaagaattctaccagaagataaaaaatgaaaaggttcaacatgaatagtag 1617
 QY 2214 ccagaagaagtggaaggttaccagaagaagtggaagaggtaccagaagaagtggaagag 2273
 Db 1618 gtgagaagaattctaccagaagaagataaaaaatgaaaaggttcaacatgaatagtag 1677
 QY 2274 gtaccagaagaagtggaagaagtggaagaagtagaagaagtagaggtaccagcggtagta 2333
 Db 1678 gtgagaagaattctaccagaagaagataaaaaatgaaaaggttcaacatgaatagtagag 1737
 QY 2334 gaagtagaagtagcagcggtagtagaagaagaaggtgcccagaagaagtagaagaagaaga 2393
 Db 1738 gaagaattctaccagaagaagataaaaaatgaaaaggttcaacatgaatagtagaggtgaa 1797
 QY 2394 gaagaggaagaaccagtagaggaagaagtagtattacaattagataaccatcggaaga 2453
 Db 1798 gaattcttcagaagaattgtgaattggaagaagtaccatcacaacaaatacaaatgaa 1857
 QY 2454 gatatacaattagacaacaaacagaaga 2482
 Db 1858 aattgaaactataaaaccagagaagaa 1886

RESULT 22

AAT78868

ID AAT78868 standard; cDNA; 5361 BP.

AC AAT78868;

AT 08-OCT-1997 (first entry)

DE P. falciparum liver stage antigen-3 coding sequence.

XX Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum;
 KW prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic;
 KW glycosyl-phosphatidylinositol membrane anchoring sequence; antibody;
 KW vaccine; immunotherapy; malaria; ds.
 XX Plasmodium falciparum.
 OS WO9641877-A2.
 PN 27-DEC-1996.
 PD 12-JUN-1996; 96WO-FR00894.
 XX 13-JUN-1995; 95FR-0007007.
 XX (INSP) INST PASTEUR.

XX Daubersies P, Druilhé P;
 XX WPI: 1997-065464/06.
 DR P-PSDB; AAW24790.
 XX Plasmodium falciparum poly:peptide(s) and related nucleic acids -
 PT derived from the liver stage antigen-3, useful for malaria vaccine
 PT prodn. and diagnosis
 XX Claim 20; Fig 2A-I; 69pp; French.
 PS This sequence corresponds to the coding sequence for a Plasmodium
 XX falciparum strain K1 pre-erythrocytic liver stage antigen-3 (LSA-3)
 CC protein. The gene sequence was isolated by screening a P. falciparum
 CC strain T9/96 library with serum from a missionary treated by prophylaxis
 CC (for strain T6/96 see FR9101286). Of 20 clones isolated, clone 729S was
 CC used to screen a library generated from Thai strain K1. One clone
 CC contained a 6.85 kb insert including the genomic sequence AAT78867.
 CC The gene comprises a 1.8 kb region encoding 3 major blocks of 4 amino
 CC acid repeats and a 3' hydrophobic region corresponding to a
 CC glycosyl-phosphatidylinositol membrane anchoring sequence. The
 CC invention relates to new polypeptides of at least 10 amino acids derived
 CC from the LSA-3 polypeptide with the exception of the peptides AAW24791-4.
 CC The LSA-3 peptides can be used to raise antibodies and as vaccines for
 CC immunotherapy of malaria.
 XX Sequence 5361 BP; 2388 A; 431 C; 1169 G; 1373 T; 0 other;
 SQ
 Query Match 5.1%; Score 170.2; DB 18; Length 5361;
 Best Local Similarity 44.1%; Pred. NO. 4.9e-19;
 Matches 911; Conservative 0; Mismatches 1138; Indels 18; Gaps 4;
 QY 407 tgaagaagtcacaaacggttagatgaggtgaatcatcattagagaggtttccactgaaga 466
 Db 306 tgaagaagtcacaaacggttagatgaggtgaatcatcattagagaggtttccactgaaga 365
 QY 467 tatggaaacagagagatggctcaacagatgatcggaacagagaagaggaactactcgtga 526
 Db 366 ttagatgataataaaaaatttagaagagccgaagatataaagaaaatacttattaaag 425
 QY 527 tatggaagagagaagaagagcgtggcgatattggaagcagggggaagagcgtggtattgga 586
 Db 426 taatataagaagaacaaagaaaattatttgacaattttattataataattggacaaa 485
 QY 587 agcagggggaagaactggcgatttggaagcagggggaagaacactggcgattttggaagcag 646
 Db 486 ttcagaacaaacagagaagtgatcagaataatgacaagtcagtgatgaaactttttaaaga 545
 QY 647 ggaagaagcgtggtgatttggaagcagggggaagaacactggcgattttggaagcaggggaaga 706
 Db 546 attattaaatagtgtagatgtaattggaagagtaaaagaaaataattttggaggaagatca 605
 QY 707 aactggagatgcggaacactgaagaaggaagcaactggagatgcggaacactgaagaatggagc 766
 Db 606 agttaatgacgatatttttaataagtttagtaaaaggtttcaacagaagaacacacaaa 665
 QY 767 aactgtgtatgtagacacagaagatagttcagctgtagtgagcagagcaaaaaatcatattcc 826
 Db 666 tgttgaagaaaaagttgaagaaggtgtagaagaaaatgacgaagaagtgtagaagaaaa 725
 QY 827 tgcctcaagaaaatgtacaacctgccgatagtaataatgacccctcttttggagatatttggga 886
 Db 726 tgtagaagaaaatgtagaa-----gaaaatgacgcaggaagtgtagcctcaagtggtga 779
 QY 887 taaagataataatttttgatcoataattaaagatttcgagccactatttcgacaaatattggtgc 946
 Db 780 agaaagtagcttcaagttgtgatgaaagtatagattcaagtagttgaagaaaatgtagc 839
 QY 947 gggtagctgctaaacatgttacgggacagaagaattgcacaatgaaacctgtaccattaccagt 1006
 Db 840 tccaactgttgaagaaatcgtagctccaaagtgt---tgtagaagtggtgctccaagtgt 896

Qy	1007	ggcagaagaagccgcgcgaagtaccagcggaagaaatagatgccactccagaggatgacctt	1066
Db	897	tgaagaagtgtagaagaataattggaagaagtgtagctgaaatgttgagaagaatgt	956
Qy	1067	cgattagatttacagaatctcccgaggaagtagaattagattatagaagaagcaac	1126
Db	957	agctgaaaaattgtagaagaagtgtagctgaaaaattgagaagaagtgtagctgaaaaatgt	1016
Qy	1127	tgaagaagaatacaacggaagtgggaccaaaggaagaagcacaacccgaagaatttagatgc	1186
Db	1017	tgaagaataatcgtagctccaactgtgagaataatcgtagctccaactgttgagaataatgt	1076
Qy	1187	cactcagaggaatggatttcgcattagacgaaactgcgaagggagaacacgagaagaacgt	1246
Db	1077	agctcccaagtgtgtgtagaagtgtagctccaagtgttgagaagaagtgtagagaagaatgt	1136
Qy	1247	agagggaagaagaacagaagaagctgcagaaggaagaagatcataagaagaactccagaag	1306
Db	1137	tgaagaagtgtagctccaactgttgagaagaatcgtagctgaaaatgttgagaagaatgt	1196
Qy	1307	agaagaagaattgagggcaactccagaggaatgatttcgattagatggaactacattaga	1366
Db	1197	agctgaaaaattgtagaagaagtgtagctgaaaaattgtagaagaagtgtagctgaaaaatgt	1256
Qy	1367	agaaaccgaagaactgcgaagggagaagaacccgtagagggagaagaacccgtagaggg	1426
Db	1257	tgaagaataatcgtagctccaactgttgagaagaatcgtagctccaactgttgagaagaatgt	1316
Qy	1427	agaagaaccgtgagggagaagaagctgcagaaggaagaagaagtagtagaggaactcc	1486
Db	1317	agctccaagtgtgttagaagtgtagctccaagtgttgagaagaagtgtagagaagaatgt	1376
Qy	1487	agaggtagactccaattagaagaaccatcaggaagaaggaaggaagggagaaggaagaga	1546
Db	1377	tgaagaagtgtagctgaaaaattgtgagaagaagtgtagctgaaaaattgttgagaagaatgt	1436
Qy	1547	aggggaaggaagaagaagcgttagtagagcgtgccagtagtgccgaacccgttagaagt	1606
Db	1437	agctgaaaaattgtagaagaagtgtagctgaaaaattgtagaagaagtgtagctgaaaaatgt	1496
Qy	1607	---agtgactccctgcagcctgtcaaacccaattgctgtccacgcgcgaatgaaacttt	1663
Db	1497	tgaagaagtgtagctgaaaaattgtgagaagaagtgtagctgaaaaattgttgagaagaatcgt	1556
Qy	1664	attcgtgtatcttagataacgatttaacgtatgcagacattacatcccttgagccatt	1723
Db	1557	agctccaactgttgagaataatcgtagctccaactgttgagaagaattgttagctccaagt	1616
Qy	1724	atttaacaaactcctaagatccgtgatgcagagagagcgtgaacatgccatccaagga	1783
Db	1617	tgtagaagtgtagctccaagtgttgagaagaagtgtagaagaataattgttgagaagaatgt	1676
Qy	1784	agcacctgtacaagtaccagtgagcagtagggccgcgcaagaagtgcacacggaagaatt	1843
Db	1677	agctgaaaaattgtgagaagaagtgtagctgaaaaattgtgagaagaagtgtagctgaaaaatgt	1736
Qy	1844	gatgcactccaagggacgatttcgaaattagaagaagaactgcagaagctccagaggaag	1903
Db	1737	tgaagaagtgtagctgaaaaattgtgagaagaatcgtagctccaactgttgagaagaatcgt	1796
Qy	1904	agaattagattagaaggaaggaagaaccaacggaagaagaagcagaagaagaaggaagcc	1963
Db	1797	agctccaactgttgagaataatcgtagctccaagtgttaga-----aagtgtagctcc	1850
Qy	1964	aacagaagggagagtcaggaagaagaattagaggaactccagaggaacttcgattcgaat	2023
Db	1851	aagtgtagaagaagtgtagaagaagaattgtgagaagaagtgtagctgaaaaattgttaga	1910
Qy	2024	agaagaaccacagggaagaagtagtagaagaacccgttagggcggaagaagaactgcagaag	2083
Db	1911	aagctgtagctgaaaaattgtgagaagaagtgtagctgaaaaattgttagaagaatcgttagctcc	1970

Qy	2084	agaagaagtgaagaaggtacctgcagaagtgaagaagtggaagaaggtaccctgcagaagt	2144		
Db	1971	aactgttgaagaatactgtagctccaactgttgaagaataattgtagctccaagtgttgtaga	2030		
Qy	2144	agaagaagtgaagaaggtaccagaagaagtgaagaaggtaccgcagaagtagaagaagt	2203		
Db	2031	aagtgtgctccaagtgttgaagaagaagtgaagaagaataattgaagaagaagtgtagctga	2090		
Qy	2204	ggaagaaggtaccagaagaagtgaagaaggtaccagaagaagtgaagaaggtaccagaaga	2263		
Db	2091	aaatgttgaagaaggtagctgaaataattgaagaagaagttagctgaaaaatgttgaaga	2150		
Qy	2264	agtgagaagaagtaccagaagaagtggaagaagtggaagaagtggaagaagtagaagaagtagaggtacc	2323		
Db	2151	aagtgtagctgaaataattgtgaagaatactgtagctccaactgttgaagaatactgtagctcc	2210		
Qy	2324	agcgtgtagtagaagtgaaggtaccagcggttagtagaagaagaaggtgccagaagaagttaga	2383		
Db	2211	aactgttgaagaataattgtagctccaagttgttagaagaagttggctccaagttgtgaaga	2270		
Qy	2384	agaagaagaagaagaaggaagaaccagtagagaagaagaagatgtattacaattagtaataacc	2443		
Db	2271	aagtgtagaagaataattgtgaagaagaagttagctgaaataattgtgaagaagaagttagctga	2330		
Qy	2444	atcggagaagaatatacaattagacaa	2470		
Db	2331	aaatgttgaagaagaagtagctgagctgaaaa	2357		
RESULT 23					
AAT78867					
ID	AAT78867 standard; DNA; 6152 BP.				
XX	AAT78867;				
AC					
XX					
DT	08-OCT-1997 (first entry)				
DE	P. falciparum liver stage antigen-3 genomic sequence.				
KW	Plasmodium falciparum; pre-erythrocyte; liver stage antigen; serum; prophylaxis; Thai strain; gene organisation; exon; intron; hydrophobic glycosyl-phosphatidylinositol membrane anchoring sequence; antibody; vaccine; immunotherapy; malaria; ds.				
OS	Plasmodium falciparum.				
Key	Location/Qualifiers				
FT	77..5605				
CDS	/*tag= a				
FT	/product= liver stage antigen-3				
FT	77..274				
FT	/*tag= b				
FT	/number= 1				
FT	275..442				
FT	/*tag= c				
FT	/number= 1				
FT	443..5602				
FT	/*tag= d				
FT	/number= 2				
PN	W09641877-A2.				
XX					
PD	27-DEC-1996.				
XX					
PF	12-JUN-1996; 96WO-FR00894.				
XX					
PR	13-JUN-1995; 95FR-0007007.				
XX					
PA	(INSP) INST PASTEUR.				
XX					
PI	Daubersies P, Druilhé P;				
XX					
DR	WPI; 1997-065464/06.				

QY 2144 agaagaagtgaagaggtaccagaagaagttagaagaggtaccgcagaaagttagaagaagt 2203
Db 2275 aagtgtgctcccaagtgttgaagaagttagaagaataatttgaagaagttagctga 2334
QY 2204 ggaagaggtaccagaagaagtgaagaggtaccagaagaagtgaagaggtaccagaaga 2263
Db 2335 aaatgttgaagaagttagctgaaatgttgaagaagttagctgaaatgttgaaga 2394
QY 2264 agtgaagaggtaccagaagaagtgaagaagtgaagaagtgaagaagttagctgaa 2323
Db 2395 aagtgtagctgaaatgttgaagaatacttagctccaaactgttgaagaatacttagctcc 2454
QY 2324 agcgttagtgaagtagaagtaccagcgttagtagaagaagaaggtgaccagaagaagttaga 2383
Db 2455 aactgttgaagaatacttagctccaaactgtttagaagaagtgtgctccaaactgttgaaga 2514
QY 2384 agaagaagaagaaggaagaaccagtagaggaagaagatgtattacaattagtaataacc 2443
Db 2515 aagtgtgaagaatactgttgaagaagttagctgaaatgttgaagaagttagctga 2574
QY 2444 atcggaagaagatatatacaattagacaa 2470
Db 2575 aaatgttgaagaagttagctgaaaa 2601
RESULT 24
AAZ23891
ID AAZ23891 standard; DNA: 49999 BP.
XX
AC AAZ23891;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO genomic DNA fragment 1.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aligner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS Example 3; Page 69-97; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

Query Match 5.1%; Score 169.4; DB 20; Length 49999;
Best Local Similarity 66.2%; Pred. No. 9.5e-19;
Matches 260; Conservative 0; Mismatches 131; Indels 2; Gaps 1;
QY 2034 acaggagaagaagttagaagaaccgttagagggcggaagaactgcagaaggaagaagt 2093
Db 7380 atagaagaagaagaagagggagggagggaggaagaagaagaagaagaagaagaag 7439
QY 2094 gaagaagtcctcacaagaatagaagaagtgaagaaggtacctcagaagttagaagaagt 2153
Db 7440 gaagaagaaggaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 7499
QY 2154 gaagaggtaccagaagaagttagaagaggtaccgcagaaagttagaagaagttagaagaagt 2213
Db 7500 gaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 7559
QY 2214 ccgagaagtggaagaggtaccagaagaagttagaagaggtaccagaagaagttagaaga - 2272
Db 7560 gaagaagacagagagagggagggagggaggggggggaggaagaagaagaagaagaaga 7619
QY 2273 -ggtaccagaagaagtgaagaagtgaagaagttagaagaagttagaagaagttagaaga 2331
Db 7620 ggaagcggagagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 7679
QY 2332 tagaagtagaagtcaccagcggtagtagaagaagaaggtgccagaagaagttagaagaagaag 2391
Db 7680 gagaaggaagaagaagggagggagggaggaagaagaagaagaagaagaagaagaagaag 7739
QY 2392 aagaagaagaagaaccagtagtagaagaagaagt 2424
Db 7740 aagaagaagaagaagagggagggagggagggagg 7772
RESULT 25
AAZ23896
ID AAZ23896 standard; DNA: 49999 BP.
XX
AC AAZ23896;
XX
DT 25-JAN-2000 (first entry)
XX
DE Murine LOBO homologue genomic DNA fragment 2.
XX
KW LOBO; long bones; bone development; bone extension; skull; osteopathic;
KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
KW spondyloepiphyseal dysplasia; achondroplasia; murine; ds.
XX
OS Mus musculus.
XX
PN WO9950284-A2.
XX
PD 07-OCT-1999.
XX
PF 26-MAR-1999; 99WO-EP02055.
XX
PR 27-MAR-1998; 98DE-1013799.
XX
PA (ROSE/) ROSENTHAL A.
XX
PI Rosenthal A, Rump A, Hess J, Aligner T, Wirth T;
XX
DR WPI; 1999-601320/51.
XX
PT Nucleic acids encoding proteins which influence bone development,
PT useful for treating and studying bone disorders -
XX
PS Example 3; Page 161-189; 391pp; German.
XX
CC This invention describes novel nucleic acids (I; designated LOBO (long
CC bones)) encoding proteins influencing bone development in mammals. The
CC proteins of the invention reduce and/or inactivate bone extension (i.e.
CC development), with exception of the skull and have osteopathic activity.
CC The nucleic acid molecules, proteins and antibodies can be used in
CC diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
CC and nucleic acid molecules, etc. are useful for production of transgenic
CC animals, especially a transgenic mouse for the study of diseases
CC associated with bone development, e.g. spondyloepiphyseal dysplasia and
CC achondroplasia. This sequence encodes the murine LOBO protein described
CC in the method of the invention.
XX
SQ Sequence 49999 BP; 13210 A; 11814 C; 10825 G; 14150 T; 0 other;

QY 2025 gaagaacacacagaggaagaagtagtagaagaacccgtagaggcggaagaacactgcagaagga 2084
 Db 538 atagaagaacacacagtagtagcggaagaagtagtcgtcgagaagaacacccaatagtggaagaa 597
 QY 2085 gaagaagtggaagaggtaccctgcagaagtagtagaagaagtggaagaggtaccctgcagaagta 2144
 Db 598 gatgtagtagaagaacactaatgttagtgagaagtagtagaagaagtagtagaagaacacccagtagtgaa 657
 QY 2145 gaagaagtggaagaggtaccagaagaagtagtagaagaggtaccctgcagaagtagtagaagaagt- 2203
 Db 658 gaagaatgatagaagaacactcctgttagttgaagaagtagtagaagaacacccagtagtt 717
 QY 2204 -ggaagaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtaccagaa 2261
 Db 718 gaagaagaagtttagaagaacacccagtagtggaagaagaatagttagaagaacaccca 777
 QY 2262 gaagtgaagaggtaccagaagaagtggaagaagtagtagaagaagtagtagaagaagtagtaggta 2321
 Db 778 gtagtgaagaagtagtagtagaagaacacccagtagtagtggaagaagaagtagtgtagaaga 837
 QY 2322 ccagcgtagtagaagaagtagtagtagtagtagtagaagaagaggtgccagagaagta 2381
 Db 838 acaccagtagtagaagaagtagtagtagtagaagaacacccagtagtagtagaagaagaatgata 897
 QY 2382 gaagaagaagaagaaggaagaacccagtagtaggaagaagaagtagtagtagaagaatgata 2441
 Db 898 gaagaacacccagtagtagtagaagaagaagtagtagaagaacacccagtagtagtggaagaagaa 957
 QY 2442 ccacgcgaagaagatatataattagacaaacacaaagaa 2479
 Db 958 gttgtggaagaacacacagtaattgaagacgcccggaa 995

RESULT 29

ID AAA50254

XX AAA50254 standard; DNA; 1926 BP.

AC AAA50254;

XX 07-NOV-2000 (first entry)

DE Epstein Barr virus nuclear antigen 1 protein (EBNA1) DNA.

XX EBV; nuclear antigen 1; EBNA1; episome; transfection; selection;

KW gene therapy; ds.

XX Epstein-barr virus.

XX WO200047778-A1.

XX 17-AUG-2000.

XX 11-FEB-2000; 2000WO-US03547.

XX 11-FEB-1999; 99US-0249585.

XX (PHAR-) PHARMACOEPIA INC.

XX Horlick RA, Chelsky D;

XX WPI; 2000-515062/46.

XX P-PSDB; AAY95856.

XX Stably transfecting eukaryotic cells with at least one episome for the

XX production of a desired protein in vitro and for gene therapy -

XX Disclosure; Fig 2; 53pp; English.

XX The present sequence is that of DNA encoding the Epstein-Barr virus

XX (EBV) nuclear antigen protein 1 (EBNA1, see AAY95856). EBNA1 is

XX utilised in a novel method for obtaining a eukaryotic cell that is

XX stably transfected with at least one episome. This method involves

XX transfecting a eukaryotic cell with: (1) a first episome comprising

CC an EBV origin of replication (oriP, see AAA50253), a gene encoding a
 CC first protein whose expression results in cell death and a
 CC selectable marker for eukaryotic cells; and (2) a second episome
 CC comprising an EBV oriP and a gene encoding a second protein, where
 CC expression of the second protein prohibits the occurrence of cell
 CC death resulting from expression of the first protein to produce a
 CC doubly transfected cells which also express an antigen that
 CC promotes retention of the episomes by the cells. The doubly
 CC transfected cells are maintained under conditions in which the
 CC first and second proteins and the selectable marker are expressed,
 CC and the selective pressure specified by the marker is maintained.
 CC Under these conditions, only cells containing both episomes live.
 CC Preferably, EBNA1 is expressed from 1 of the episomes, and the
 CC protein of interest from the other episome. Either or both
 CC epitopes may further comprise a nucleic acid sequence encoding a
 CC protein desired to be expressed in the cell (e.g. a therapeutic
 CC protein), a nucleic acid encoding an RNA that is not intended to
 CC be translated (e.g. a therapeutic RNA), or a DNA sequence used as
 CC a tag for the cells. The method is applicable to cell culture or
 CC intact organisms, for gene therapy. It allows the rapid
 CC establishment of eukaryotic cells that stably and reliably express
 CC a gene of interest, using a novel method of selection, and
 CC maintenance of that selection without the need for exogenous
 CC selection factors, such as antibiotics.

XX Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

SQ

Query Match

Best Local Similarity 5.0%; Score 167.2; DB 21; Length 1926;

Matches 373; Conservative 0; Mismatches 343; Indels 0; Gaps 0;

QY 1739 caaggatcctgatgcagagaggtgtaacagttaccatcaacaaaggaagcctgtacaagt 1798
 Db 246 caaaggagccacggtggaacagagcagagagcagagggcaggaagcagagagg 305
 QY 1799 accagtggcagtagggcccgcgcaagaagtgcacacggaagaattgatgcactccaaga 1858
 Db 306 ggcagagcagggaggggcagagcagagagggcagggcagggcagggcagggcaggg 365
 QY 1859 ggaacgtttcgaattagaagaactgcagaagctccagaggaaggaagattagattaga 1918
 Db 366 aggggcagagcagggaggggcagagcagagagggcagggcagggcagggcaggggc 425
 QY 1919 aggaagaagtagaaccacgaagaagagccaaagagagagagccacgaagaagagaagt 1978
 Db 426 aggagcagagaggggcagagcagagagagggcagggcagggcagggcagggcaggg 485
 QY 1979 gccagaagaagaattagaggcaactccagagagcgatttcgaattagaagaaccacagg 2038
 Db 486 ggcaggggggcaggggggcagagcagagaggggcagggcagggcagggcaggggcag 545
 QY 2039 agaagaagtagaagaaccgtagagggcgaaagaactgcagaaggaagaagaagtggaaga 2098
 Db 546 aggggcagagcagggaggggcagaggggggcagagggggcagggcagggcaggggcag 605
 QY 2099 ggtacctgcagaagtagaagaagtggaaaggttcacctgcagaagtagaagaagtggaaga 2158
 Db 606 aggagcagaggggcagaggggcagaggggcagaggggcagaggggcagggcagggcag 665
 QY 2159 ggtaccagaagaagtagaagaggtaccctgcagaagtagaagaagtggaagaggtaccaga 2218
 Db 666 aggggcaggggcaggggcagaggggcagagggggcagggcagggcagggcagggcaggg 725
 QY 2219 agaagtggaaaggtaccagaagaagtggaagaggtaccagaagaagtggaagaaggtacc 2278
 Db 726 agggaggggcagggcaggggggcagaggggggcagggcagggcagggcagggcagggcag 785
 QY 2279 agaagaagtggaagaagtggaagaagtagaagaagtagaggtaccagcggttagtagaagt 2338
 Db 786 agcagagggggcaggggcagaggggcagagcagagggaggggcaggggcagggcagggag 845
 QY 2339 agaagtaccagcggttagtagaagaagaaggtgccagaagaagtagaagaagaagaaga 2398

```

Db      846 aggggcaggaggcaggagcaggaggcaggaggcaggaggcaggaggcaggagg 905
QY      2399 ggaagaaccagtagaagaagaagatgtattacaattagtaataaccatcggaagaag 2454
Db      906 ggcaggagcaggaggcaggaggcaggaggcaggaggcaggaggcaggaggag 961

RESULT 30
AAF82902
ID      AAF82902 standard; DNA; 1926 BP.
XX
AC      AAF82902;
XX
XX      29-JUN-2001 (first entry)
XX
DE      EBV tethering protein EBNA1 encoding DNA.
XX
KW      Histone H1; tethering protein; LANA; gene therapy; multiple sclerosis;
KW      Parkinson's disease; Huntington disease; diabetes; human herpesvirus 8;
KW      EBV; latency-associated nuclear antigen; LANA; EBNA1; ds.
XX
OS      Epstein-barr virus.
XX
FH      Key
FT      CDS
FT      1..1926
FT      /*tag= a
XX
XX      W0200125484-A2.
XX
PD      12-APR-2001.
XX
XX      29-SEP-2000; 2000WO-US26908.
XX
PR      01-OCT-1999; 99US-0410399.
XX
XX      (UNMI ) UNIV MICHIGAN.
XX
XX      Robertson ES, Cotter MA;
XX
XX      WPI: 2001-281736/29.
XX      P-PSDB; AAB62332.
XX
PT      A composition for use in gene therapy comprises an expression vector
PT      that includes a nucleic acid sequence encoding a nucleic acid binding
PT      protein
XX
PS      Disclosure; Fig 9C; 60pp; English.
XX
CC      The invention provides a composition comprising nucleic acid, histone H1
CC      protein and expression vector operatively encoding a protein suitable
CC      for tethering the nucleic acid to the histone H1 protein, where the
CC      tethering protein is LANA. The composition is useful in aiding the
CC      retention of the viral DNA in the host cell. The viral vector encodes a
CC      protein suitable for tethering DNA to Histone H1. Methods for screening
CC      for compounds which are agonistic or antagonistic for the tethering of
CC      viral proteins to histone H1 and DNA binding sites are useful for
CC      developing the method of viral transfer. The composition has applications
CC      to gene therapy, including the treatment of multiple sclerosis,
CC      Parkinson's disease, Huntington disease and diabetes. The present
CC      sequence represents the nucleotide sequence of the Epstein-barr virus
CC      (EBV) tethering protein ENNA1.
XX
XX      Sequence 1926 BP; 487 A; 352 C; 872 G; 215 T; 0 other;

Query Match      5.0%; Score 167.2; DB 22; Length 1926;
Best Local Similarity 52.1%; Pred. No. 1.3e-18;
Matches 373; Conservative 0; Mismatches 343; Indels 0; Gaps 0;

QY      1739 caaggatccctgacgagagagcgtgtacaagtagtaccatcaaaaggaagcactgtacaagt 1798
Db      246 caaaggaccaccagggtgtgaacagagcaggagcaggagcggaggcaggagcaggag 305

```

```

QY      1799 accagtggcagtagggcccgcaagaagtgccaacggaagaattgatgcaactccaaga 1858
Db      306 ggcaggagcaggaggaggcaggagcaggagcaggaggcaggaggcaggaggcaggagg 365
QY      1859 ggcagatttcgaattagaaggaaactcagaagctccagaggaaggaggaattagattataga 1918
Db      366 aggggcaggagcaggaggaggcaggagcaggagcaggaggcaggaggcaggaggaggc 425
QY      1919 aggagaaggagaaccaacggaagaagcgaagaagagagagagagagagagagagagaagt 1978
Db      426 aggagcaggaggaggaggcaggagcaggaggaggaggaggaggaggaggaggaggaggagg 485
QY      1979 gccagaagaagaattagaggcaactccagaggcgagcatttcgaatttagaagaaccacaagg 2038
Db      486 ggcaggaggggcaggaggggcaggagcaggaggaggaggaggaggaggaggaggaggagg 545
QY      2039 agaagaagtagaagaacccgtagaggggcgaagaactgcagaaggaggaagaagtgaaga 2098
Db      546 aggggcaggagcaggaggaggaggcaggaggggcaggaggggcaggaggcaggaggaggagg 605
QY      2099 ggtacctgcagaagtagaagaagtggaagaggtacctgcagaagtgaagaagaagtgaaga 2158
Db      606 aggagcaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 665
QY      2159 ggtaccagaagaagttagaagaggtaccgcagagaagtagaagaagtggaaaggtaccaga 2218
Db      666 aggggcaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 725
QY      2219 agaagtggaaagaggtaccagaagaagtggaaagagggtaccagaagaagtggaaaggtacc 2278
Db      726 aggaggggcaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 785
QY      2279 agaagaagtggaaagaagtggaaagaagtagaagaagtagaggtaccagcggtagtagaagt 2338
Db      786 agcaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 845
QY      2339 agaagtaccagcggttagaagaagaaggtgccagaagaagtagaagaagaagaagaaga 2398
Db      846 aggggcaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 905
QY      2399 ggaagaaccagtagagaagaagaagatgtattacaattagtaataaccatcggaagaag 2454
Db      906 ggcaggagcaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggaggagg 961

```

Search completed: June 14, 2002, 15:37:24
Job time: 15436 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 10:11:49 ; Search time 4066.1 Seconds
(without alignments)
17174.160 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagaataca.....tctataataataataattc 3337

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797556 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pi.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pi.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description

1	3337	100.0	3337	6	ARI51085
2	3337	100.0	3337	6	I23337
3	342.8	10.3	2069	23	E10125
4	342.8	10.3	3399	23	E10126
5	330.6	9.9	11554	8	AP002460
6	318.6	9.5	81019	8	AB073159
7	307.8	9.2	144505	2	AL662884
8	307.2	9.2	1902	8	AF413055
9	304	9.1	215046	2	AC011767
10	299	9.0	172307	2	AC044842
11	294.6	8.8	142277	2	AC016204
12	283.6	8.5	16377	3	AF202180
13	283.6	8.5	102195	2	AC091338
14	278.8	8.4	272545	2	AC090533
15	278.4	8.3	174241	9	AC018634
16	277.4	8.3	282611	2	AL645746
17	273.8	8.2	165077	10	AC084382
18	268	8.0	1686	23	E08995
19	263.8	7.9	4630	3	PF111A
20	256.2	7.7	5420	3	AF056936
21	255.8	7.7	174019	2	AC027267
22	253.2	7.6	170276	2	AC106251
23	252.2	7.6	199508	2	AC098555
24	252	7.6	13297	10	AF311727
25	247.2	7.4	179149	9	AC009785
26	244.8	7.3	157158	2	AC098094
27	242	7.3	132449	9	AL365272
28	242	7.3	178022	2	AC018348
29	242	7.3	179337	2	AC091304
30	241.8	7.2	206136	9	AC068037
31	240	7.2	184855	2	AL606744
32	238.8	7.2	185497	2	AC073446
33	237.2	7.1	159971	2	AC091951
34	236.2	7.1	22398	5	FRU271723
35	235.6	7.1	208006	2	AC087877
36	234.8	7.0	3720	14	S76368
37	234.8	7.0	43658	14	HSV3PRGEN
38	234.8	7.0	112930	14	HSGEND
39	232.8	7.0	214864	9	AC069544
40	231.4	6.9	180385	9	AC007461
41	231.4	6.9	230839	10	AC068561
42	230	6.9	227605	2	AC073564
43	228	6.8	12029	3	AE001373
44	227.4	6.8	185994	2	AC002042
45	226	6.8	163542	2	AC106201

ALIGNMENTS

RESULT 1	ARI51085	ARI51085	Sequence 1 from patent US 6231861.	3337 bp	DNA	linear	PAT 08-AUG-2001
LOCUS	ARI51085	Sequence 1	from patent US 6231861.				
DEFINITION	ARI51085	Sequence 1	from patent US 6231861.				
ACCESSION	ARI51085	Sequence 1	from patent US 6231861.				
VERSION	ARI51085.1	GI:151117135					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unknown.						
REFERENCE	1 (bases 1 to 3337)						
AUTHORS	Barnwell,J.W.						
TITLE	Plasmodium vivax blood stage antigens, antibodies, and diagnostic assays						
JOURNAL	Patent: US 6231861-A 1 15-MAY-2001;						
FEATURES	Location/Qualifiers						
source	1..3337						
BASE COUNT	1304 a 467 c 875 g 691 t						
ORIGIN							
Query Match	100.0%	Score 3337;	DB 6;	Length 3337;			

Best Local Similarity 100.0%; Pred. No. 0;									
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
QY	1	gaattccggtaaagtaaacactatggttcgtatctatataataaccttactaattttatc	60						
Db	1	GAATTCGGTAAAGTAACAACTATGTGTTTCGTATCTATATAACCTTACTAATTTATC	60						
QY	61	ttttgctttcttttaattcatcttcaacagtaagataaaaaataatctataaaaactgc	120						
Db	61	TTTTTGCTTTCTTTTAATTCATGCTTCAACAGTAAAGATAAAATAATCTATAAAACTGC	120						
QY	121	tatatatacatatattcataagtggttcattgtgaatgcgatcttataatttacgta	180						
Db	121	TATATATACATATATATTCATAAGTGGCATTTGTGAATTGCGATCATTTAAATTTACGTA	180						
QY	181	aaacaataattgaaaaaatt	240						
Db	181	AAAAAATAATTGAAAAAATT	240						
QY	241	aattgaaaatgcttctgatgtttagaggtgagatccctcaaacgcggttttag	300						
Db	241	AATTGAAAAATGCTTCTGATGATGTTGTAGAGGTGCGAGATCCTTCAAACGCGGTTTAG	300						
QY	301	aattagaagagaaaatttttgatgagaattcaagttgatgatgaactcttttagatgcta	360						
Db	301	AATTAGAAGAGGAAAAATTTTGATGGAATTCAGGTGATGATGAACCTCTTTTAGATGCTA	360						
QY	361	ccccgaagatgacttgccttaacagatttccaattgaagacgatgaggaagtcaacg	420						
Db	361	CCCCGAGATGACTTTGCTTTAACAGATTTGCCAATTTGAACAGATGAGGAAGTCAACG	420						
QY	421	aaactgttagtgaggtgaatcaattaggagaggtttccactgaagatatggaacagaag	480						
Db	421	AAACGTTAGATGAGGTGAATCATATTAGGAGAGGTTTCCACTGAAGATATGGAACAGAAAG	480						
QY	481	atgctcaacagatgatacgaaaacagaagaagactaactcgtgatatggaaggagaag	540						
Db	481	ATGGCTCAACAGATGATACGGAACAGAAAGAGGACTACTGGTGATATGGAAGGAGAAAG	540						
QY	541	aagaagctggcgtatggaagcagggggaagaagcttggtgatttggaaagcaggggaagaa	600						
Db	541	AACAAGCTGCGATATGGAAGCAGGGGAAGAAGCTGGTGNATTGGAAGCAGGGGGAAGAA	600						
QY	601	ctggcatttggaaagcaggggaagaactggcgatttggaaagcaggggaagaagctggcg	660						
Db	601	CTGGCATTTTGGAAAGCAGGGGAAGAACTGGCGATTTTGGAAAGCAGGGGAAGAGCTGGTG	660						
QY	661	atttgaagcaggggaagaaactgcgatttgaagcaggggaagaaactgaagatgcg	720						
Db	661	ATTTGGAGCAGGGGAAGAAACTTGGCGATTTTGGAAAGCAGGGGAAGAACTTGGAGATGCGG	720						
QY	721	aaactgaagaagcgaactggagatgcggaactgaaatggagcaactgtgatgtag	780						
Db	721	AAACTGAAGAAGCAGCAACTGGAGATGCGGAAACTGAAATGGAGCAACTGTGTATGTAG	780						
QY	781	acacagaagatagttcagctgatggagcagaaaaagtaactgttctcgtctcaagaagaaatg	840						
Db	781	ACACAGAAGATAGTTCAGCTGATGGAGCAGAAAAAGTACATGTTCTGTCTCAAGAAAAATG	840						
QY	841	tacaactcgcgatagtaaatgacccctcttggaaagtatttggataaagataataatt	900						
Db	841	TACAACCTGCCGATAGTAAATGATGCGCCTCTTTGGAAAGTATTTTGGATAAAGATATAATTT	900						
QY	901	ttgatcataatgaagatttcagccactattcgaacaaaatttggcgggttactgctaaac	960						
Db	901	TTGATCATATTAAGATTTTCAGGCCACTATTTCGAACAAATTTGTGGCGGTACTGCTAAAC	960						
QY	961	atgttcagggaagaagtgcgaatgaacactgtacattaccagttggcagaagaccgcg	1020						
Db	961	ATGTTACGGGACAGAATTTGCCAATGAAACCTGTACCATTTACCAGTTGGCAGAAAGACCCG	1020						
QY	1021	cgcaagtaccagcggaagattagatgccactccagagatgaacttcgattagatgta	1080						
Db	1021	TACCTCCAGATAGAAAGTGGAAAGAGTACCTCCAGAAAGTAGAAGAGTGGAAAGG	1080						
Db	1021	CGCAAGTACCACGCGAAGAAATTAGATGCCACTCCAGAGGATGACTTCGATTAGATGTTA	1080						
QY	1081	cagaactctccagagaagtagaattagtagtaaaagcaactgaagaagaatcaa	1140						
Db	1081	CAGAATCTCCGAGGAAGTAGAATTAGTATTAGTAGAAGGCACTGAAGAAGATCAA	1140						
QY	1141	cgaagtgggaccacgcggaagaggaaccacccaagaataattagatgccactccagagatg	1200						
Db	1141	CGGAAGTGGACCAACGGAAGAGGACCAACCAAGAAATTAGATGCCACTCCAGAGGATG	1200						
QY	1201	gatttcgacttagacgaactcagaagagagaagaacacgtagaggagagaagaaa	1260						
Db	1201	GATTTTCGCAATTAGCAAACTTCAGAAAGGAGAAACAGAAACGTTAGGGGAGAAAGAAA	1260						
QY	1261	cagaagaactcagaagagagaatcagaagaactccagaagagagaagagagttag	1320						
Db	1261	CAGAAGAACTCGAAGGAGGAAGTATCAGNAGAACTCCAGNAGGAAGAAAGAGATTAG	1320						
QY	1321	aggcaactccagagatgatttcgcattagatggaaactacattagagaacccgaagaaa	1380						
Db	1321	AGGCAACTCCAGAGGATGATTTTCGCATTAGATGGAACCTACATTAGAAGAAACCGAAGAAA	1380						
QY	1381	ctgcagaagagaagaacacgtagagggagaagaacccgtagaggagagaagaaacgtag	1440						
Db	1381	CTGCAGAAGGAGAAGAAACCCGTAGAGGGAGAAACCCGTAGAGGGAGAAAGAACCGTAG	1440						
QY	1441	agggagaagaagctgcagaagagagaagagttagaggcaactccagagagatgacttcc	1500						
Db	1441	AGGGAGAACAAGCTGCAGAAGGAGAGAAGAGCTTACAGGCAACTCCAGAGGATGACTTCC	1500						
QY	1501	attagaagaacacacaggaagaggaaggggaagagaaggaaggggaaggaag	1560						
Db	1501	AATTAGAAGAACCATCAGGAGAGGAGAAAGGGGAGAGAGAGAGAGGAGGAGAGAAAG	1560						
QY	1561	gagaagcgttagtagcagtcagtagtgccgaaccgtagaagtagtgcctcctgcctc	1620						
Db	1561	GAGAAGCGTTAGTAGCAGTGCAGTAGTGCCTGCGCAACCGGTAGAAGTAGTACTTCTGCTC	1620						
QY	1621	agcctgtcaaaccaatggctcctcaacgcgagatgaaactttattcgttgatatcttag	1680						
Db	1621	AGCCTGTCAAAACAATGGTCTCCAAACGCAGATCAAACTTTATTCTGTTGATATCTTAG	1680						
QY	1681	ataagcatttaacgtagtcagacattacatcctttgagccattatttaacaaactccta	1740						
Db	1681	ATAACGATTTAACGATGTCAGACATTTACATCTCTTTTCAGGCCATTTATTAACAACTCTCA	1740						
QY	1741	aggatcctgatgcagagagagctgtaacagtaccatcaaggaagcacctgtacaagtac	1800						
Db	1741	AGGATCTCTGATGCAGGAGAGGCTGTAAACAGTACCATCAAGGAAAGCACCTGTACAAGTAC	1800						
QY	1801	cagtgagtagggccgcgcaagaagtgccaacggaagaattgatacgaactccaagagg	1860						
Db	1801	CAGTGGCAGTAGGGCCGCGCAAGAAAGTGCCAACGGAAGAAATTGATGCAACTCCAAGAGG	1860						
QY	1861	agcattcgaatttagaagaaactcagaagcctccagaagaggaagattagattagaag	1920						
Db	1861	AGCATTTTCAATTAGAAGGAACCTGCAAGAGCTCCAGGAAGGAGAAATTAGTATTAGAAG	1920						
QY	1921	gagaaggaagaaacacgcggaagagagccagaagagagagccaaacagaggaagagtg	1980						
Db	1921	GAGAAGGAGAACCAACGGAAGAGAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGTGC	1980						
QY	1981	cagaagaagaattagaggcaactccagaggaagcatttcgaatttagaagaaccaacaggag	2040						
Db	1981	CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTTCGAATTTAGAAAGAACCAACAGAGG	2040						
QY	2041	aagaagtaagaagaacccgtagaggcggaagaactcaagaagagagaagtggaagagg	2100						
Db	2041	AGAAGTAGAAGAAACCGTAGAGGGCGAAGAACTGCAGAAGGAGGAAGTAGTGGAAAGAGG	2100						
QY	2101	tacctcagaagtagaagaagtggaaaggtacctcagaagttagaagaagtggaaagagg	2160						
Db	2101	TACCTCCAGATAGAAAGTGGAAAGAGTACCTCCAGAAAGTAGAAGAGTGGAAAGG	2160						


```
Qy 2761 cagaatatagtcataattcaacaaaggaaagaccagctggtttgagaaacagagagt 2820
Db 2761 CAGATATAGTCATAATTCAACAAAAGAAAGACCAGCTGGTTGAGAAAACGAGAGT 2820
Qy 2821 ggagcagcagagaatggaaaaatggtttaagcagaagtcaaatcccaattgattcac 2880
Db 2821 GGAGCAGCAGAGAATGGAAAAATGTTTAAAGCAGAACTCAAAATCCCAATTTGATTCAC 2880
Qy 2881 actgaaaaatgtagaacgacactcattccaattatttaaaattctgtgaaagata 2940
Db 2881 ACTTGAAAAATGATGAACGACACTCATTCCATTTATTAAATCTGTGTGAAGATA 2940
Qy 2941 tgtcacatttgaacaaagaaacaaagaaatggttaatgaatcactcgaaagaaacg 3000
Db 2941 TGTCACAATTTGAAAAACAAGAAACCAAGAAATGGTTAATGAATCACTGGAAGAACG 3000
Qy 3001 aacggggttatggttcgaatcatttgaagtattgaccacatacaaatatttaattgtg 3060
Db 3001 ACGGGGTTATGGTTCTGAATCATTGTAAGTTATGACCACATCAAAATTTATTAAATGTGG 3060
Qy 3061 ctaagagtcgagaatggtaccgtgccaatcctaataataatagagaagaaagaaactca 3120
Db 3061 CTAAGAGTCGAGATGGTACCGTGCCCAATCCTTAATATAATAGAGAAAGAACTCA 3120
Qy 3121 tgaattggtttctctaaagaaacgaataatttaggacaaagaaatggaaaaatggact 3180
Db 3121 TGAATGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 3180
Qy 3181 cattgaaaaaagttaaaattttttgttcaattcaattcaattgtgtacaacattttctgaaa 3240
Db 3181 CATTTGAAAAAAGTTAAATTTTGTGTTCAATTCATTTGTTCAACATTTTCTTGAAAA 3240
Qy 3241 cgcttaaccaaggaagaatggaatcaattgttgaatgaataaaagtgtgaattatagaa 3300
Db 3241 CGCTTAACCAAGGAATGGAATCAATTTGTTAATGAATATAAAGCTTTGAATTATAGAA 3300
Qy 3301 aaagacagattattctctataaataaataatttc 3337
Db 3301 AAAAACAAGATTATTCTCTTATAAATAAATAATTC 3337

RESULT 3
E10125 standard; RNA; UNC; 2069 BP.
AC E10125;
SV E10125.1
XX 08-OCT-1997 (Rel. 52, Created)
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)
XX DNA encoding an immunogenicity protein.
KW JP 1995284392-A/1.
XX unclassified
OS unclassified.
RN [1]
RP 1-2069
RA Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;
RT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZON IMMUNOGENIC PROTEIN AND GENE
RT RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZON";
RL Patent number JP1995284392-A/1, 31-OCT-1995.
RL DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST.THE.
XX
CC OS Leucocytozon caulleryi
CC FN JP 1995284392-A/1
CC PD 31-OCT-1995
CC CF 19-APR-1994 JP 1994080643
CC PI DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO
CC PC C12N15/09,A61K39/015,C12P21/02;
```

```
CC CC strandedness: Double;
CC CC topology: Linear;
CC FH Key Location/Qualifiers
CC FH source 1..2069
CC FT /organism="Leucocytozon caulleryi"
CC FT /strain="shizuoka"
CC FT misc_feature 1..25
CC FT note="EcoRI adaptor"
CC FT misc_feature 2045..2069
CC FT note="EcoRI adaptor"
CC FT misc_feature 25..2046
CC FT product="immunogenicity protein"
XX FH Key Location/Qualifiers
FH FH source 1..2069
FH FT /db_xref="taxon:32644"
FH FT /organism="unidentified"
XX SQ Sequence 2069 BP; 1216 A; 167 C; 443 G; 243 T; 0 other;

Query Match 10.3%; Score 342.8; DB 23; Length 2069;
Best Local Similarity 51.8%; Pred. No. 2.7e-40;
Matches 874; Conservative 0; Mismatches 807; Indels 7; Gaps 4;

Qy 955 ctaaacatgttacgggacagaattgccaatgaaacctgtaccattaccagtggcagaag 1014
Db 383 CAACATTTATGAAGAGAAATAAATTTATATAGTAGATTATTTACCAGAAATTAATGTAG 442
Qy 1015 agccgcgcgaagtaccagcggaagaatttagatgccactccagaggtgactctgcattag 1074
Db 443 AAGAATCAGATGAACAGAACATATAACATATGAATAGATAATGACATACAGAAGAGC 502
Qy 1075 atgttacagaattcccgagggaagtagaattagta ctagatgaaggaggcaactgaagaag 1134
Db 503 ATGAAAAAGTAAACATGATGAAGAGAAAAAGAAAGTAAACATGATGAAGAAATAGAAAAAG 562
Qy 1135 aatcaacggaagtgggacgaagagaccacgaagaccacgaagaattagatgccactccag 1194
Db 563 AAGAGCATGAAGAAGTAAATACATGAAGAGAAAAAGAAAGTAAACATGATGAAGAATAG 622
Qy 1195 aggatggatttc--gcattagacgaactgcagaagagagaacagagaacgtta--gag 1251
Db 623 AAAAAAGACAGCATGAAGAAGTAAATACATGAAGAGAAAAAGAAAGTAAACATGAAAG 682
Qy 1252 gagagaaacagaagaagctgcagaaggaagagatcagagaagaactccagagagggaag 1311
Db 683 AAAAAAGAAAAAGAGCATGAAGAAGTAAATACATGAAGAGAAAAAGAAAGTAAACAC 742
Qy 1312 aagagttagaggaactccagagagatttcgcatttagatggaactacattagaagaaa 1371
Db 743 ATGAAGAAATAGAAAAAGAGAGCATGAAGAAGTAAATACATGAAGAGAAAAAGAAAG 802
Qy 1372 ccgaagaactgcagaaggaagaaacccgttagagggagagaagaaacccgttagagggaag 1431
Db 803 TAACATGTAAGAAATAGAAAAAGAGAGCATGAAGAAGTAAATACATGAGGAGAAAAAG 862
Qy 1432 aaacgttagagggaagaagctgcagaaggaagaagaagatttagaggcaactccagaag 1491
Db 863 AAGAAGTAAACATGAAGAAGTAAAGAAAAAGAGAGAGCATGAAGAAGTAAATACATGAAG 922
Qy 1492 atgacttccaatttagaagaaccatcaggaaggaaggaaggaaggaaggaaggaaggaag 1551
Db 923 AAAAGAAAGAGTAAACATGAAGAATAGAAAAAGAGAGCATGAAGAAGTAAATACATG 982
Qy 1552 aaggagaaggaagcgttagtagcagtgccagtagtgccgaacccggttagaagtagtga 1611
Db 983 AAGAAGAAAAAGAAAGTAAACATGAAGAAAAAGAAAAAGAAAGAGCATGAAGAAGTAA 1042
Qy 1612 ctctgctcagcctgtcacaaccaatggtcgctccaacgagcagatgaaactttattcgtg 1671
```


Qy	2149	aagtggaaagagggtaccagaagaagtagaaggtaccgcagaaagttagaagaagtcggaag	2208
Db	2732	AAAGAAGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG	2791
Qy	2209	aggtaccagaagaagtcggaagaggtaccagaagaagtcggaagaggtaccagaagaagtcg	2268
Db	2792	AAGAAGAAGAAGATGAGGAAGAGAGAGAAGAGAAGAAAATCGGGAAGAGAAGAAG	2851
Qy	2269	aagaaggtaccagaagaagtcggaagaagtcggaagaagtagaagaagtagagggtaccagcgg	2328
Db	2852	AAAAAAGGAAGAAGAAAGAAAGAAAAAGACAGCATGAAGAAGAAGCTAAACACATGAAG	2911
Qy	2329	tagtagaagtagaaggtaccagcggtagtagaagaagaggtgcacagaagaagtagaagaag	2388
Db	2912	AAGNAGAAGAAAAAGTAAACATGAGAGAGAGAGAAAAAGTAAACATGAGAGAGAGAAA	2971
Qy	2389	aagaagaaggaaggaaccagtagaggaagaagatgtattacaattagtagtaataccatcgg	2448
Db	2972	ATGTTAAACATATGAAGAAGAAAGAAAAAGTAAACATGAAGAAGAAGAAAAAGTAAAC	3031
Qy	2449	aagaagatacacaattagacaaaccaagaagaagcgaattaggcctcgtgaaatttatcta	2508
Db	3032	ATGAAGAAGAAGAAAAAGTAAACATGAAGAAGAAGAAAATGTAACATATGAAGAAGAAG	3099
Qy	2509	tcatacgacatgcactaccaagacggtccaaagggaatttaggaagaagaagaagaacatg	2568
Db	3092	AGAAAAAAGTAAACACATGAAGAAGAAGAAAAAGTAAAT-GAAGAAGAAGAAGATAATG	3155
Qy	2569	cagtgtagtccatgaaaccagaagaatttgcgaagaagaattcacaatctcagaatggc	2628
Db	3151	AAAGTATAAAAAANAANAANAANAANAANAANAAGGGGGGGGGCCATGTCACCCGGATCC	3210
Qy	2629	tcacattc	2636
Db	3211	TCGAATTTC	3218

RESULT	5				
AP002460					
LOCUS	AP002460	11154 bp	DNA	linear	PLN 27-DEC-2000
DEFINITION	Arabidopsis thaliana genomic DNA, chromosome 3, BAC clone:FD9.				

Accession	Source	Organism	Gene	Protein	Accession	Source	Organism	Gene	Protein
	Arabidopsis thaliana (strain:Columbia) DNA, clone:FID9.	Arabidopsis thaliana	Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.						

STRUCTURAL ANALYSIS OF ARABIDOPSIS THALIANA CHROMOSOME 3. III
UNPUBLISHED
2 (BASES 1 TO 111554)
REFERENCE

TITLE Direct Submission
JOURNAL Submitted (06-JUN-2000) Yasukazu Nakamura, Kazusa DNA Research Institute, Department of Plant Gene Research; 1532-3, Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: ynakamu@kazusa.or.jp, ynakamu@kazusa.or.jp)

COMMENT
Address for correspondence: kaos@kazusa.or.jp
For the latest information on annotation of this clone, please see
http://www.kazusa.or.jp/kaos/cgi-bin/aqd_graph.cgi?c=FLD9

'product' or 'note' qualifiers. Genes that have no significant protein similarity are described as 'unknown protein'.

(Informatics Group, Oak Ridge National Laboratory,
<http://compbio.ornl.gov/Grail-1.3/>).

NetGene2 (S.M. Hebsgaard, et al., CBS, Technical University of Denmark, <http://www.cbs.dtu.dk/services/NetGene2/>) and

SplicePredictor (Volker Brendel, Stanford University, <http://gremlini.zool.iastate.edu/cgi-bin/sp.cgi>).
Genes encoding tRNAs are predicted by tRNAScan-SE (Sean Eddy, Washington University School of Medicine, St. Louis, <http://genome.wustl.edu/eddy/tRNAScan-SE/>).
This sequence may not be the entire insert of this clone. It may be shorter because we remove overlaps between neighboring submissions.
The 5' clone is T89 and the 3' clone is T7B9.
Location/Qualifiers
1. .111554
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/chromosome="3"
/clone="F1D9"
/clone_lib="IGF BAC"
complement(join(4747. .5195,5297. .5754,5827. .6284, 6457. .6803,6864. .7079,7216. .7286,7326. .7810,7914. .8253, 8311. .8737,9083. .9347))
/note="gene_id:F1D9.1"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="helicase-like protein"
join(11589. .11736,11836. .12019,12305. .12497,12774. .12896, 13021. .13080)
/note="gene_id:F1D9.2"
/codon_start=1
/evidence=not_experimental
/product="ATP-dependent CLPB protein-like"
/protein_id="BAA97091.1"
/db_xref="GI:8777573"
/translation="MLEQILHERVIAQDLDESADAIKRCAGISDPNRPITAFMFM GPTSVVQLVGASVGVGDEGKSTEVVRRPYSVVOQDEIEKPHPNVFSILLQLLD DGRITDHSGLHILIRNNEIDKEAVVEMKQOVVELARXFKPKFNNRIDEIVS QPLNSSVISKIVELQVQMRQMKRLEQPNKLEITLVNLKGFDAENFILLDAQPF NKLVIKMNNAHVKN"
complement(join(19018. .19216,19317. .19390,19600. .19746))
/note="contains similarity to unknown protein
gb|AAC97228.1
gene_id:F1D9.3"
/codon_start=1
/evidence=not_experimental
/protein_id="BAA97092.1"
/db_xref="GI:8777574"
/translation="MAKERHRVGVGVREEDVKRMKLIMPYLNESHQOLRALFSGYP STLKITNYWKELEENKHTKQEFIAFSRSGFSGAYIPMCAKIINKDFGKHGLAVL PGKIITLFTLVSKSKLGIKDKLIIVIVIVKLVMP"
join(21817. .22004,22075. .22143)
/note="contains similarity to ATP-dependent CLPB protein
gene_id:F1D9.4"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(22489. .23812)
/note="contains similarity to retroelement pol
polyprotein
gene_id:F1D9.5"
/codon_start=1
/pseudo
/evidence=not_experimental
join(24404. .24543,24659. .24879)
/note="contains similarity to ATP-dependent CLPB protein
gene_id:F1D9.6"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(25073. .29379)
/note="gene_id:F1D9.7"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="retroelement pol polyprotein-like"

CDS
join(30097. .30190,30265. .30551,30614. .31510,31758. .32036)
/note="gene_id:F1D9.8"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="ATP-dependent CLPB protein-like"
join(40639. .40982,41779. .42326,42393. .42794,42903. .42938, 43020. .43142,43226. .43333,43425. .43481,43588. .43650, 43733. .43973,44026. .44107,44503. .44674,44813. .44940, 45043. .45198)
/note="gb|AAD25852.1
gene_id:F1D9.9
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(46889. .47780,47984. .48332,48695. .48783)
/note="gene_id:F1D9.10
pir||T01475
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
join(51111. .51234,51470. .52398,52474. .52672,52762. .53360, 53433. .53558,53717. .53809,53908. .54010,54110. .54210, 54472. .54531,54627. .54728,54795. .55047,55142. .55226, 55334. .56637,57320. .57389,57601. .58187)
/note="gene_id:F1D9.11
pir||T02599
strong similarity to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(join(58782. .58995,59066. .59150,59231. .59498, 60276. .61922,61970. .62259,62305. .62482))
/note="gene_id:F1D9.12"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="maize transposon MUDR protein-like"
complement(join(63866. .64008,64160. .64226,64346. .64414, 64504. .64620,64743. .64790,64956. .65053,65183. .65264, 65486. .65566,65677. .65718,65854. .65952,66106. .66144))
/note="contains similarity to human diaphanous 1
gene_id:F1D9.13"
/codon_start=1
/evidence=not_experimental
/protein_id="BAB01984.1"
/db_xref="GI:9294133"
/translation="MLSKVKIPLPDLMSVLADESVIDVDQVNLKFCPTKEAEAL LKGFIGNKETIGRCQFFLELLKVPVETKLVFKIQFHSQVSGSKLRKIMOTIL SILGNALNHGTARGSAIGFHLDSLLKLTDTSRNSKMTLMHLYLCKGVLAEPGLNFP KDMIQLYLAEMQATSKGLEKVQVEFTASETDQCIKSHFMNLEFSLVAGEVRS L ASLYSTVGSADALALYFGEDPARVPFEQVSTLQNFVRIEVRSHSEENCKQVEFEKRR AQKEAENEKIKKGVYNE"
complement(66524. .66974)
/note="gb|AAD23008.1
gene_id:F1D9.14
similar to unknown protein"
/codon_start=1
/pseudo
/evidence=not_experimental
complement(join(68701. .69561,69697. .70591,70640. .70667, 71526. .72709,72898. .73300,73370. .73533,73687. .74594))
/note="gene_id:F1D9.15"
/codon_start=1
/pseudo
/evidence=not_experimental
/product="retroelement pol polyprotein-like"
complement(75178. .76058)
/note="gb|AAD23008.1
gene_id:F1D9.16
similar to unknown protein"


```
misc_feature 53828..58528
/Note="similarity to fragments of Tma retrotransposons"
LTR 58529..59847
misc_feature 59848..62072
/Note="similarity to fragments of Athila retrotransposons"
LTR 62073..62855
LTR 65457..65959
misc_feature 65960..70568
/Note="similarity to fragments of Tat retrotransposons"
repeat_region 70569..71234
/rpt_family="(GAA)n"
misc_feature 71235..73234
/Note="similarity to fragments of Athila retrotransposons"
LTR 73235..75775
LTR 75776..77556
CDS join(78124..78212,78321..78572,78639..78807,79136..79851,
79967..80048)
/codon_start=1
/evidence=not_experimental
/product="putative bZIP transcriptional activator-like
protein"
/protein_id="BAB83611.1"
/db_xref="GI:18149198"
/translation="MQLICDIDGTIYLYKAPKKVILLIYRFLVLTLSKNVSPAIH
PFISQISLQNIIEFYNIIFTYIPILLHOQIPHLIKHEIKDKFGQNLTKSGHK
PTLKLTAHEHRLPALLNPASLLTSATKRNNLTDSFIRRRQOSIFPSPFLVSIQIS
SPSRFTINEMASSKGSQNRNLGCKQKALLPSPFTGGTFSADFVSPVSIQIS
VQKLGEGNAHHRTSSEFLIEQPSWLDLNEPETPVKRGHRRSSDSDFAYVDVP
VGFDVDTLWDGGRYNNNGFSNHRVPKESDYLRSOPVPEYPSAHLKQKIRPMDSL
PDGARGPNSSCLESSTLRSCSGSLRDTKAVSAADSKDFINFAKSFERDND
PLAKSATSEADYKRAQQAQFAQSRVRKIQYIAELERNVQLQV"
BASE COUNT 24352 a 16182 c 17018 g 23467 t
ORIGIN

Query Match 9.5%; Score 318.6; DB 8; Length 81019;
Best Local Similarity 67.3%; Pred. No. 7e-37;
Matches 450; Conservative 0; Mismatches 219; Indels 0; Gaps 0;

Qy 1822 aagaagtcacaggaaggaattgatgcacactccaagaggagcatttcgaattagaagaa 1881
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70570 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70629

Qy 1882 ctgcagaagctccaggaaggaaggaattattagtagaaggaaggaaggaaggaag 1941
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70630 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70689

Qy 1942 aagagccaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 2001
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70690 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70749

Qy 2002 ctccagaggacgatttcgaattagaagaaacacagaggaaggaaggaaggaagcctag 2061
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70750 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70809

Qy 2062 aggcgcagaagaaactgcagaaggaaggaaggaaggaaggaaggaaggaaggaag 2121
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70810 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70869

Qy 2122 tgggaagggctacctgcagaaggaaggaaggaaggaaggaaggaaggaaggaag 2181
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70870 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70929

Qy 2182 taccgcagagtagaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 2241
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70930 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 70989

Qy 2242 aagtcgaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaaggaag 2301
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 70990 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71049

Qy 2302 aagtagaaggaagtagaggtaccagcggttagtagaagtagaagtagaagtagaag 2361
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```
Db 71050 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71109
Qy 2362 aagaggtgcagagaagtagaagaagaagaagaagaagaccagtagtagaagaag 2421
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71110 AAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71169
Qy 2422 atgtattacaattagtaataaccatcggaagaagataatacaattagacaacaaagaa 2481
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71170 CAGAAGAAGAAGAAGCAGAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAG 71229
Qy 2482 acgaattag 2490
||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71230 AAGAATTAG 71238

RESULT 7
AL662884/c 144505 bp DNA linear HTG 08-FEB-2002
LOCUS Homo sapiens chromosome 6 clone Xbac-300A18, *** SEQUENCING IN
DEFINITION PROGRESS ***, 9 unordered pieces.
ACCESSION AL662884
VERSION AL662884.4 GI:18643821
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
TITLES Sims, S.
JOURNAL Direct Submission
COMMENT Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi:18151543.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: bpg300A18
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 140263 bases at least Q40
Consensus quality: 141340 bases at least Q30
Consensus quality: 142357 bases at least Q20
Insert size: 143705; sum-of-contigs
Insert size: 144041; 6.8% error; agarose-fp
Quality coverage: 9.03x in Q20 bases; sum-of-contigs Quality
coverage: 9.45x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 6934 7033: gap of 100 bp
* 7034 28893: contig of 21860 bp in length
* 28894 28993: gap of 100 bp
* 28994 31651: contig of 2658 bp in length
* 31652 31751: gap of 100 bp
* 31752 45643: contig of 13892 bp in length
* 45644 45743: gap of 100 bp
* 45744 61827: contig of 16084 bp in length
* 61828 61927: gap of 100 bp
* 61928 107959: contig of 46032 bp in length
* 107960 108059: gap of 100 bp
```


* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 23438: contig of 23438 bp in length
* 23439 23538: gap of 100 bp
* 23539 24096: contig of 558 bp in length
* 24097 24196: gap of 100 bp
* 24197 24746: contig of 550 bp in length
* 24747 24846: gap of 100 bp
* 24847 25309: contig of 463 bp in length
* 25310 25409: gap of 100 bp
* 25410 26051: contig of 642 bp in length
* 26052 26151: gap of 100 bp
* 26152 27047: contig of 896 bp in length
* 27048 27147: gap of 100 bp
* 27148 27754: contig of 607 bp in length
* 27755 27854: gap of 100 bp
* 27855 28543: contig of 689 bp in length
* 28544 28643: gap of 100 bp
* 28644 29290: contig of 647 bp in length
* 29291 29390: gap of 100 bp
* 29391 30194: contig of 804 bp in length
* 30195 30294: gap of 100 bp
* 30295 30697: contig of 403 bp in length
* 30698 30797: gap of 100 bp
* 30798 31446: contig of 649 bp in length
* 31447 31546: gap of 100 bp
* 31547 32368: contig of 822 bp in length
* 32369 32468: gap of 100 bp
* 32469 33461: contig of 993 bp in length
* 33462 33561: gap of 100 bp
* 33562 34177: contig of 616 bp in length
* 34178 34277: gap of 100 bp
* 34278 35035: contig of 758 bp in length
* 35036 35135: gap of 100 bp
* 35136 35892: contig of 757 bp in length
* 35893 35992: gap of 100 bp
* 35993 36855: contig of 863 bp in length
* 36856 36955: gap of 100 bp
* 36956 37837: contig of 882 bp in length
* 37838 37937: gap of 100 bp
* 37938 38976: contig of 1039 bp in length
* 38977 39076: gap of 100 bp
* 39077 39933: contig of 857 bp in length
* 39934 40033: gap of 100 bp
* 40034 41139: contig of 1106 bp in length
* 41140 41239: gap of 100 bp
* 41240 42285: contig of 1046 bp in length
* 42286 42385: gap of 100 bp
* 42386 43508: contig of 1123 bp in length
* 43509 43608: gap of 100 bp
* 43609 44697: contig of 1089 bp in length
* 44698 44797: gap of 100 bp
* 44798 46002: contig of 1205 bp in length
* 46003 46102: gap of 100 bp
* 46103 47134: contig of 1032 bp in length
* 47135 47234: gap of 100 bp
* 47235 48609: contig of 1375 bp in length
* 48610 48709: gap of 100 bp
* 48710 50120: contig of 1411 bp in length
* 50121 50220: gap of 100 bp
* 50221 51074: contig of 854 bp in length
* 51075 51174: gap of 100 bp
* 51175 52266: contig of 1092 bp in length
* 52267 52366: gap of 100 bp
* 52367 53467: contig of 1101 bp in length
* 53468 53567: gap of 100 bp
* 53568 54859: contig of 1292 bp in length
* 54860 54959: gap of 100 bp
* 54960 56493: contig of 1534 bp in length
* 56494 56593: gap of 100 bp
* 56594 58097: contig of 1504 bp in length

* 58098 58197: gap of 100 bp
* 58198 60019: contig of 1822 bp in length
* 60020 60119: gap of 100 bp
* 60120 61572: contig of 1453 bp in length
* 61573 61672: gap of 100 bp
* 61673 63511: contig of 1839 bp in length
* 63512 63611: gap of 100 bp
* 63612 64527: contig of 916 bp in length
* 64528 64627: gap of 100 bp
* 64628 66684: contig of 2057 bp in length
* 66685 66784: gap of 100 bp
* 66785 68609: contig of 1825 bp in length
* 68610 68709: gap of 100 bp
* 68710 70596: contig of 1887 bp in length
* 70597 70696: gap of 100 bp
* 70697 72728: contig of 2032 bp in length
* 72729 72828: gap of 100 bp
* 72829 76282: contig of 3454 bp in length
* 76283 76382: gap of 100 bp
* 76383 78231: contig of 1849 bp in length
* 78232 78331: gap of 100 bp
* 78332 84930: contig of 6599 bp in length
* 84931 85030: gap of 100 bp
* 85031 90279: contig of 5249 bp in length
* 90280 90379: gap of 100 bp
* 90380 101025: contig of 10646 bp in length
* 101026 101125: gap of 100 bp
* 101126 109913: contig of 8788 bp in length
* 109914 110013: gap of 100 bp
* 110014 120982: contig of 10969 bp in length
* 120983 121082: gap of 100 bp
* 121083 132369: contig of 11287 bp in length
* 132370 132469: gap of 100 bp
* 132470 151295: contig of 18826 bp in length
* 151296 151395: gap of 100 bp
* 151396 177432: contig of 26037 bp in length
* 177433 177532: gap of 100 bp
* 177533 215046: contig of 37514 bp in length.

FEATURES

source

Location/Qualifiers

1. .215046
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="15"
/map="15"
/clone="Rp11-26F2"
/clone_lib="RPCI-11 Human Male BAC"
1. .23438
/note="assembly_fragment"
clone_end:SP6
vector_side:left
23539..24096
/note="assembly_fragment"
24197..24746
/note="assembly_fragment"
24847..25309
/note="assembly_fragment"
25410..26051
/note="assembly_fragment"
26152..27047
/note="assembly_fragment"
27148..27754
/note="assembly_fragment"
27855..28543
/note="assembly_fragment"
28644..29290
/note="assembly_fragment"
29391..30194

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

Query Match 9.1%; Score 304; DB 2; Length 215046;
Best Local Similarity 47.5%; Pred. NO. 8.6e-35;
Matches 969; Conservative 0; Mismatches 1065; Indels 6; Gaps 2;

QY 399 gaagacgatgagggaagtcaacgaacgttagatgaggtgaattcattaggagaggtttcc 458

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

1 (bases 1 to 172307)
Biren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 2, clone RP11-178E20
Unpublished
2 (bases 1 to 172307)

Biren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
Collimore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczeky,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Menus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Testaye,S., Theodore,J., Tirrell,A., Travers,M., Triggilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (12-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 25, 2000 this sequence version replaced gi:7543808.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

TITLE
JOURNAL
COMMENT

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9002
Center clone name: 178_E20
----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156589 bases at least Q40
Consensus quality: 164477 bases at least Q30
Consensus quality: 167863 bases at least Q20
Insert size: 171000; agarose-fp
Insert size: 170107; sum-of-contigs
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.7 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 23 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 19: contig of 19 bp in length
* 20 119: gap of 100 bp
* 120 1921: contig of 1802 bp in length
* 1922 2021: gap of 100 bp
* 2022 3689: contig of 1668 bp in length
* 3690 3789: gap of 100 bp
* 3790 6531: contig of 2742 bp in length
* 6532 6631: gap of 100 bp
* 6632 8134: contig of 1503 bp in length
* 8135 8234: gap of 100 bp

* 8235 11919: contig of 3685 bp in length
* 11920 12019: gap of 100 bp
* 12020 15244: contig of 3225 bp in length
* 15245 15344: gap of 100 bp
* 15345 17344: contig of 2000 bp in length
* 17345 17444: gap of 100 bp
* 17445 21830: contig of 4386 bp in length
* 21831 21930: gap of 100 bp
* 21931 25081: contig of 3151 bp in length
* 25082 25181: gap of 100 bp
* 25182 29736: contig of 4555 bp in length
* 29737 29836: gap of 100 bp
* 29837 36853: contig of 7017 bp in length
* 36854 36953: gap of 100 bp
* 36954 41270: contig of 4317 bp in length
* 41271 41370: gap of 100 bp
* 41371 46765: contig of 5395 bp in length
* 46766 46865: gap of 100 bp
* 46866 57601: contig of 10736 bp in length
* 57602 57701: gap of 100 bp
* 57702 67879: contig of 10178 bp in length
* 67880 67979: gap of 100 bp
* 67980 78335: contig of 10256 bp in length
* 78336 78335: gap of 100 bp
* 78336 89177: contig of 10841 bp in length
* 89177 89276: gap of 100 bp
* 89277 106081: contig of 16805 bp in length
* 106082 106181: gap of 100 bp
* 106182 121387: contig of 15206 bp in length
* 121388 121487: gap of 100 bp
* 121488 137207: contig of 15720 bp in length
* 137208 137307: gap of 100 bp
* 137308 153745: contig of 16438 bp in length
* 153746 153845: gap of 100 bp
* 153846 172307: contig of 18462 bp in length.

FEATURES
Source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-178E20"
/clone_lib="RPC1-11 Human Male BAC"
1. 19
/note="assembly_fragment
vector_side:left"
120. 1921
/note="assembly_fragment"
2022. 3689
/note="assembly_fragment"
3790. 6531
/note="assembly_fragment"
6632. 8134
/note="assembly_fragment"
8235. 11919
/note="assembly_fragment"
12020. 15244
/note="assembly_fragment"
15345. 17344
/note="assembly_fragment
clone_end:SP6
vector_side:right"
17445. 21830
/note="assembly_fragment"
21931. 25081
/note="assembly_fragment"
25182. 29736
/note="assembly_fragment"
29837. 36853
/note="assembly_fragment"
36954. 41270
/note="assembly_fragment"

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature

misc_feature


```
Db 39782 TAGCTTGCTAGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGCGAGGAGTAGCAG 39841
QY 2074 ctgcagaaggaagaagtggaagagctacctgcagaagtagaagaagtggaagaggtac 2133
Db 39842 CATCACAGTTCAGTAGCTGTGCGAGGAGGAGTAGCAGCATCAGCAGTTGCGAGTCGCTG 39901
QY 2134 ctgcagaagtagaagaagtggaagaggtaccagaagaagtagaagaaggtaccgcagaag 2193
Db 39902 TGCACAGGAGGAGTAGCAGCAGCAGCAGTTCGAGTAGCTGTGCGAGGAGTAGCAGCATCAG 39961
QY 2194 tagaagaagtggaagaaggtaccagaagaagtggaagaggtaccagaagaagtggaagaag 2253
Db 39962 TAGTTCAGTAGCTGTGCGAGGAGGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGCGCAG 40021
QY 2254 taccagaagaagtggaagaggtaccagaagaagtggaagaagtggaagaagtagaagaag 2313
Db 40022 GAGGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGCGAGGAGGAGTAGCAGCATCAGCAG 40081
QY 2314 tagaggtaccagcgttagtagaagtagaaggtaccagcgttagtagaagaagaggggtgccag 2373
Db 40082 TTGCAGTAGCTGTGCGAGGAGGAGTAGCAGCATCAGCAGTTGCGAGTAGCTGTGCGCAGGAG 40141
QY 2374 aagaagtagaagaagaagaagaagaggaaggaaggaagtagaagaagaagtagattacaat 2433
Db 40142 TAGCAGCATCAGCAGTTGCGAGCAGCTGTGCGAGGAGGAGTAGCAGCATCAGCAGTTGCGAG 40201
QY 2434 tagtaataccatcgggaagaag 2454
Db 40202 CAGCTGTGCGAGGAGGAGTAG 40222

RESULT 11
AC016204
LOCUS
DEFINITION Homo sapiens clone RP11-25C1, *** SEQUENCING IN PROGRESS ***, 5
unordered pieces.
ACCESSION AC016204
VERSION AC016204.6 GI:18644852
KEYWORDS HTG; HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 142277)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome, clone RP11-25C1
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 142277)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baldwin,J., Barna,N., Beckerly,R., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Castle,A., Collangelo,M., Collins,S., Collumore,A.,
Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Trafaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
Wyman,B., Ye,W.J., Zimmer,A. and Zody,M.
Direct Submission
JOURNAL Submitted (23-NOV-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT On Feb 11, 2002 this sequence version replaced gi:14209777.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
```

```
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L4539
Center clone name: 25_C_1
-----
```

* NOTE: This is a 'working draft' sequence. It currently consists of 5 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence, as soon as it is available and the accession number will be preserved.

* 1 24412: contig of 24412 bp in length
* 24413 24512: gap of 100 bp
* 24513 46948: contig of 22436 bp in length
* 46949 47048: gap of 100 bp
* 47049 103797: contig of 56749 bp in length
* 103798 103897: gap of 100 bp
* 103898 126333: contig of 22436 bp in length
* 126334 126433: gap of 100 bp
* 126434 142277: contig of 15844 bp in length.

FEATURES

source

1..142277
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-25C1"
/clone.lib="RPC1-11 Human Male BAC"

BASE COUNT 39176 a 32698 c 33584 g 36253 t 566 others
ORIGIN

Query Match 8.8%; Score 294.6; DB 2; Length 142277;
Best Local Similarity 47.0%; Pred. No. 2e-33;
Matches 932; Conservative 0; Mismatches 1049; Indels 4; Gaps 1;

QY 399 gaagcagatgaggaagtcacaaagcttagatgaggtggaatcattaggaaggtttcc 458
Db 131759 GGAGCAGGAGGGCGAGTGTGCGGAGCAGGAGGGCGGAGTGTGCGGAGCAGGAGGGCGAGGT 131818

QY 459 actgaagatatgaaacagaagatgctcaacagatgatacgaacacagaagaagacta 518
Db 131819 GCGGAGCAGGAGGGCGAGTGTGCGGAGCAGGAGGGCGGAGTGTGCGGAGCAGGAGGAGCA 131878

QY 519 cctgtgatatggaaggaagaagaagctgctgcatatggaagcagggaggaagaagctgt 578
Db 131879 GATGGGGAGCAGGAGGAGCAGATGCAGAAAGCAGGAGGAGCAGATGCGGGAGCAGGAGGA 131938

QY 579 gatttgaagcgggggaagaactggcgaatttgggaagcgggggaagaactggcgaatttg 638
Db 131939 GCAGATGCGGAAGCAGGAGGAGCAGATGCGGGAGCAGGAGGAGCAGATGCAGAAAGCAGGA 131998

QY 639 gaagcaggggaagaagctgtgatttgaagcaggggaagaagaactggcgaatttgaagca 698
Db 131999 GGAGCAGGTGTCAAAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGATGCGGGAGCA 132058

QY 699 ggggaagaactggagatgcggaactgaagaagggcaactggagatgcggaactgaa 758
Db 132059 GGAGGAGCAGATGCGGGAGCAGGAGGAGCAGATGCTGAAGCAGAGGAGCAGCAGGAGCA 132118

QY 759 aatgagcaactgtgtatgtatgacacagaagaatagttcagctgtatgagcagaagaagta 818
Db 132119 GGAGCAGAGNNNNNNNNNACGGGGGAGCAGGAGGAGCAGATGCGGGAGCAGGAGGAGCA 132178

QY 819 catgttcctgtcagaagaaatataaacctgcgcatagtaatgatccctctttggaagt 878
Db 132179 GATGCGGGAGCAGGAGGAGCAGATGCGGGAGCAGGAGGAGCAGATGCGGGAGCAGGAGGA 132238

QY 879 atttggataaagataataattttatcatattataaagatttcagccatttcgaacaa 938
Db 132239 GCAGATGGGAGCAGGAGGAGCAGATGTTGGGAGCAGAGGAGCAATGTTGGGAGCAGAGAA 132298

QY 939 atttggcgggtactgctaacaatgttaccgggacaagaattgccaatgaacctgtacca 998

Db 132299 GGAGCAGATGGAAGCAGGAGGAGCAAAATGGGGGAGCAGGAGGACAGATGCGAAGCA 132358
QY 999 ttaccagtgcagaagaccgcgcaagtaccagcggaagaattagatccactccagag 1058
Db 132359 GGAGGACAGGTGCGAAGCAGGAGGAGCAGATGCGGAGCAGGAGCAGATGCGGAA 132418
QY 1059 gatgaacttcgattagattgttacagaatctcccgaggaagtagaatttagtattagatgaa 1118
Db 132419 GCAGGAGGACAGATCGGGAAGCAGGAGGAGCAAAATGGGGGACACAGAAGGACAAATGGG 132478
QY 1119 gaggcaactgaagaagaatacaacggaagtgggaccacgaagaagaagacccaacccgaaga 1178
Db 132479 GGAGCAGGAGGAGCAGATCGGGAGCAGGAGGAGCAAAATGGGGGAGCAGAAAGGAGCAGAT 132538
QY 1179 ttgatgtccactccagatgatgttcgatttcgatttagacgaactgcagaaggaagaacaga 1238
Db 132539 CGCGAGCAGGAGGAGCAGATGCGGGGAGCAGNAGAAAGGAGCAGATGCGGAGCAGGAGGAGC 132598
QY 1239 aga-----aacgtagagggagaagaacagagaagctgcagaagggagaagatccagaaga 1294
Db 132599 AGATGCGGAAGCAGGAGGAGCAGATCGGAAGCAGGAGGAGCAGATGGGGGAGCAGGAGG 132658
QY 1295 aactccagaaggaagaagagtagaggcaactccagaggatgatttcgattagatgg 1354
Db 132659 AGCAGGTGCGGGGAGCAGGAGGAGCAGATGGGGGAGCAGGAGGAGCAGATGCGGAAGCAGG 132718
QY 1355 aactacattagaagaacccaagaactgcagaaggaagaacacccgtagagggaagaaga 1414
Db 132719 AGAAGCAGATGCTGAAGCAGAAAGGACAGATGGGGGAGCAGGAGGAGCAGATGGGGGAGC 132778
QY 1415 aaccgtagagggagaagaacccgtagaggaggaagaagctgcagaagggagaagaagatt 1474
Db 132779 AGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGATGCGAAGCAGGAGGAGCAAAATGGGGG 132838
QY 1475 agaggaactccagagatgacttccaattagaagaacacctcaggaaggaaggaagggga 1534
Db 132839 AGCAGAAGGAGCAGATGCGGAAGCAGGAGGAGCAAAATGGGGGAGCAGGAGGAGCAAAATGG 132898
QY 1535 aggaaggaaggaagggagaaggaagaagcgttagtagcagtgccagtccagttagtgcgca 1594
Db 132899 GGGAGCAGGAGGAGCAGATCGGAAGCAGGAGCAGCAGGTGCAAAAAGCAGGAGGAGCAGA 132958
QY 1595 accggtagaagtagtagtactctctcagcctgtcaaaccaatggtcgtcccaacggcaga 1654
Db 132959 TGGCGGAAGCAGGAGGAGCAGATGCGAGCAGGAGGAGGTGCGGGGAGCAAGAGGAGCAGG 133018
QY 1655 tgaactttattcgttatctttagataataacgatttaacgtagtcagacattacatcctt 1714
Db 133019 TGGGGGAGCAGGAGGAGCAGATGAGGAGCAGGAGGAGAGATGGGGGAGCAGGAGGAGC 133078
QY 1715 tgagccattatttaacaaatcctcaggatcctgtagcaggagagcgtgttaacagttacc 1774
Db 133079 AGATGCAGAAGCAGGAGGAGCAGATGGGGGAGGAAGGAGGAGCAGATGGGGGAGCAGGAGG 133138
QY 1775 atcaaggaagaccctgtacaagtaccagtggcagttagggccgcgaagaagatgcgaac 1834
Db 133139 AGCAGATGCGGAAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGGAGGATGGGGGAGCAGG 133198
QY 1835 ggaagaattgatcaactccaagagacgatttcggaatttagaaggaactgcagaagctcc 1894
Db 133199 AGAGCAGATGCGGGGAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGGAGCAGATGGGGGAGC 133258
QY 1895 agagggaaggaatttagtattagaaggaagaaggaaggaacccaacggaagaagagccaagaga 1954
Db 133259 AGGAGGAGCAGATGAGGAAGCAGGAGGAGCAGATGGGGGAGCAGGAGGAGCAGATGGGA 133318
QY 1955 aggaagccaacagaagggagaagatgcgaagaagaatttagadgcaactccagagaagaca 2014
Db 133319 AGCAGGAGGAGCAGATGGGGGAGCAGGAGGAGCAGATGGGGGAGGAGGAGGAGCAGATGG 133378
QY 2015 tticgaattagaagaaccaacagagaagaagtagaagaacccgtagagggcaagaac 2074

Db 133379 GGGAGCAGGAGGAGCAGATGTGGGAGCAGGAGGAGCAGATGCGGAAGCAGGAGGAGCAGA 133438
QY 2075 tgcagaagcagaagaagtgaagaaggtacctgcagaagtagaagaagtgaagaagttacc 2134
Db 133439 TGGGGGAGCAGGAGGAGCAGATGAGGAATCAGGAGGAGCAGATGGGGGAGCAGGAGGAGC 133498
QY 2135 tgcagaagttagaagaagtgaagaaggtaccagaagaagtagaagaaggtaccctccgcagaagt 2194
Db 133499 AGATGCGGAAGCAGGAGGAGCAGATGGGGGAGCAGGAGGAGCAGATGAGGAAGCAGGAGG 133558
QY 2195 agaagaagtgaagaaggtaccagaagaagtggaagaaggtaccagaagaagtggaagaaggt 2254
Db 133559 ACAGATGCGGGGAGCAGGAGGAGCAGATCGGAAGCAGATGCGGAGGAGCTGCGATTCAAG 133618
QY 2255 accagaagaagtggagaaggtaccagaagaagtgggaagaagtgggaagaagtagaagaagt 2314
Db 133619 AGGAGAGGCTGTGGGATGAGTATGAGAAGATGCAGGAGGAGGAGGAGAGATCCGGGGGC 133678
QY 2315 agaggtaccagcgttagtagaagtagaagttaccagcgttagtagaagaagaagtggtgccaga 2374
Db 133679 AGGTGAGAGAAGATGGGGGAGAGAAGAGGAGGATGGGAGCAGGAGGAGGAAGATGCAGG 133738
QY 2375 aqaag 2379
Db 133739 AGGAG 133743

RESULT 12
AF202180
LOCUS AF202180 16377 bp DNA linear INV 02-APR-2001
DEFINITION Plasmodium falciparum erythrocyte membrane-associated giant protein antigen 332 gene, complete cds.
ACCESSION AF202180 AF327072 AH010491 AF323001 AF323000 AF322999
VERSION AF202180.3 GI:13508498
KEYWORDS
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 16377)
AUTHORS Shan,Z.X., Yu,X.B., Li,X.R., Ma,C.L., Fang,J.M., Lu,J.H. and X.J.
TITLE Molecular cloning and structure analysis of the Plasmodium falciparum erythrocyte membrane-associated giant protein Ag332 (Pf332) gene
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 16377)
AUTHORS Shan,Z.X., Yu,X.B., Li,X.R. and Ma,C.L.
TITLE Direct Submission
JOURNAL Submitted (05-NOV-1999) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REFERENCE 3 (bases 1 to 16377)
AUTHORS Shan,Z.X.
TITLE Direct Submission
JOURNAL Submitted (08-DEC-2000) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REMARK Sequence update by submitter
REFERENCE 4 (bases 1 to 16377)
AUTHORS Shan,Z.X.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2001) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REMARK Sequence update by submitter
REFERENCE 5 (bases 1 to 16377)
AUTHORS Shan,Z.X.
TITLE Direct Submission
JOURNAL Submitted (24-FEB-2001) Parasitology, Sun Yat-Sen University of Medical Sciences, 74 Zhongshan Road II, Guangzhou, Guangdong 510089, China
REMARK Sequence update by submitter
COMMENT On Apr 2, 2001 this sequence version replaced gi:13111670 gi:12232085 gi:13177607 gi:12719373.

```

antigen 332
1. .16377
/codon_start=1
/product="erythrocyte membrane-associated giant protein
antigen 332"
/protein_id="AA15293.1"

```

BASE COUNT	7347 a	1411 c	3730 g	3889 t	
Query Match	8.5%	Score 283.6;	DB 3;	Length 16377;	
Best Local Similarity	46.0%	Pred. No. 8.3e-32;			
Matches 1090;	Conservative	0;	Mismatches 1264;	Indels 18;	Gaps 3;
Qy	246	gaaaatgcttctgatgattgttagaggtgagagatccttcaaacgacggttagaatta	305		
Db	5347	GAATCAGTTGCTGAAGAAATAGTAGAGGAAGAAGATCTGTTGTATGACAAATAATAGAG	5406		
Qy	306	gaagagaaaaattttagatgagaattcaggtgtagatgaaactctttagatgctaccacccc	365		
Db	5407	GAACAACATCTGTTGTAGAAAGTTGTAGAACAAAGGATCAGTTATTGAAGACAGATAGTA	5466		
Qy	366	gaagatgacttgccttaacagatttcccattgaagacgtagaggagtcacagcaaacg	425		
Db	5467	GAAGAAGAAGTAATAGATAGTGAAGAAGGATCGATATGGAAGAAGAAATGTTAGAAGATGGA	5526		
Qy	426	ttagatggagtggaatttaggagaggtttccactgaagatatggaacacagaagatggc	485		
Db	5527	GTTCTAGAACAGCAATGGGTACTGATGCTTTTATGGCACAAAGGATCTGTTATTAGAA	5586		
Qy	486	tcaacagatgatacggaacacagaagaagac---tacctggtgatatggaagggaagaa	542		
Db	5587	GAAATTGAAGAAATAGTAGATAGTGAAGAAGGATCGATATGGAAGAAGAAATGTTAGAAGATGGA	5646		
Qy	543	gaagctggcgaataggaagcaggggaagaagctgggtgatttgggaagcaggggaagaact	602		
Db	5647	TCAGCTAATGAAAAAATAGTAGAGGAAGAACCATCTCGTGTGGAAGANGTTTTAGGAAAA	5706		
Qy	603	ggcgatttgaagcaggggaagaactggcgatttgaagcaggggaagaagctgggtgat	662		
Db	5707	GAAGCATTTGTTATTGAAGAAATATAGAACAGGGTCGGTAATTGTAACAAAGTTTGAAGAT	5766		

REFERENCE
AUTHORS
Rattus
 1 (bases 1 to 102195)
 Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
 Alsbrooks,S.L., Anaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
 Benton,G., Brieva,M., Blankenburg,K., Bonnin,D., Bouck,J.,
 Bowle,S., Briveva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
 Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
 Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
 Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
 Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
 Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Deigado,O.,
 Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
 Dunn-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
 Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
 Foster,P., Franco,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
 Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
 Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
 Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
 Hollins,B., Honsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
 Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jollivet,S.,
 Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
 Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
 Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
 Loulsegad,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
 Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
 Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
 Morgan,M., Morris,S., Moser,M., Neal,D., Newtonson,J., Newtonson,N.,
 Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwuo,S.,
 Ogulu,M., Okwuonu,G., Oreguine,N., Oviedo,R., Pace,A., Payton,B.,
 Peery,J., Peters,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
 Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubenok,I., Rolfe,M.,
 Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
 Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
 Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
 Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
 Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalobos,D., Vinson,R.,
 Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
 Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Woodden,S.,
 Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
 and Gibbs,R.
Direct Submission
 Unpublished
 2 (bases 1 to 102195)
 Worley,K.C.
Direct Submission
 Submitted (18-APR-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Apr 21, 2001 this sequence version replaced gi:13661904.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: TUGG
 Center clone name: CH230-1G12
 ----- Summary Statistics
 Sequencing vector: Plasmid: M77789
 Chemistry: Dye-primer Bodipy: 65% of reads
 Assembly: Dye-terminator Big Dye: 35% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 99820 bases at least Q40
 Consensus quality: 116847 bases at least Q30
 Consensus quality: 125383 bases at least Q20
 Estimated insert size: 90042; sum-of-contigs estimation
 Quality coverage: 0x in Q20 bases; agarose-gel estimation
 Quality coverage: 1.1x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 *

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 55 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 * 1
 * 3200: contig of 3199 bp in length
 * 3299: gap of unknown length
 * 3300: contig of 2586 bp in length
 * 5885: gap of unknown length
 * 5886: contig of 2677 bp in length
 * 8662: gap of unknown length
 * 8663: contig of 2723 bp in length
 * 11485: gap of unknown length
 * 11486: contig of 2248 bp in length
 * 13933: gap of unknown length
 * 13934: contig of 2583 bp in length
 * 16516: gap of unknown length
 * 16517: contig of 2975 bp in length
 * 19591: gap of unknown length
 * 19592: contig of 2320 bp in length
 * 19692: gap of unknown length
 * 22012: contig of 2432 bp in length
 * 22112: gap of unknown length
 * 24544: contig of 1734 bp in length
 * 24644: gap of unknown length
 * 26378: contig of 2905 bp in length
 * 26478: gap of unknown length
 * 29382: contig of 2661 bp in length
 * 29482: gap of unknown length
 * 32143: contig of 2238 bp in length
 * 32144: gap of unknown length
 * 34481: contig of 2338 bp in length
 * 34482: gap of unknown length
 * 36199: contig of 2338 bp in length
 * 37019: gap of unknown length
 * 38915: contig of 1896 bp in length
 * 39015: gap of unknown length
 * 39016: contig of 1976 bp in length
 * 40992: gap of unknown length
 * 41092: contig of 1823 bp in length
 * 43014: gap of unknown length
 * 43015: contig of 2513 bp in length
 * 45227: gap of unknown length
 * 45628: contig of 1555 bp in length
 * 47182: gap of unknown length
 * 47283: contig of 1197 bp in length
 * 48479: gap of unknown length
 * 48480: contig of 2768 bp in length
 * 51347: gap of unknown length
 * 51447: contig of 1384 bp in length
 * 52831: gap of unknown length
 * 52832: contig of 1269 bp in length
 * 54200: gap of unknown length
 * 54201: contig of 2114 bp in length
 * 56414: gap of unknown length
 * 56514: contig of 1274 bp in length
 * 57788: gap of unknown length
 * 57889: contig of 1432 bp in length
 * 59320: gap of unknown length
 * 59321: contig of 1138 bp in length
 * 60558: gap of unknown length
 * 60559: contig of 1897 bp in length
 * 62555: gap of unknown length
 * 62556: contig of 1366 bp in length
 * 64021: gap of unknown length
 * 64022: contig of 1546 bp in length
 * 64122: gap of unknown length
 * 65667: contig of 1837 bp in length
 * 65668: gap of unknown length
 * 67704: contig of 1536 bp in length
 * 67705: gap of unknown length
 * 69340: contig of 1307 bp in length
 * 70647: gap of unknown length

```
* 70648 70747: gap of unknown length
* 70748 71827: contig of 1080 bp in length
* 71828 71927: gap of unknown length
* 71928 73326: contig of 1399 bp in length
* 73327 73426: gap of unknown length
* 73427 74847: contig of 1421 bp in length
* 74848 76388: gap of unknown length
* 76389 76468: gap of unknown length
* 76469 77702: contig of 1234 bp in length
* 77703 77802: gap of unknown length
* 77803 79520: contig of 1718 bp in length
* 79521 79620: gap of unknown length
* 79621 81495: contig of 1875 bp in length
* 81496 81595: gap of unknown length
* 81596 82995: contig of 1400 bp in length
* 82996 83095: gap of unknown length
* 83096 84243: contig of 1148 bp in length
* 84244 84343: gap of unknown length
* 84344 85849: contig of 1506 bp in length
* 85850 85949: gap of unknown length
* 85950 87081: contig of 1132 bp in length
* 87082 87181: gap of unknown length
* 87182 88530: contig of 1349 bp in length
* 88531 88630: gap of unknown length
* 88631 89886: contig of 1256 bp in length
* 89887 89986: gap of unknown length
* 89987 91224: contig of 1238 bp in length
* 91225 91324: gap of unknown length
* 91325 92431: contig of 1107 bp in length
* 92432 92531: gap of unknown length
* 92532 93848: contig of 1317 bp in length
* 93849 93948: gap of unknown length
* 93949 95271: contig of 1323 bp in length
* 95272 95371: gap of unknown length
* 95372 96724: contig of 1353 bp in length
* 96725 96824: gap of unknown length
* 96825 98292: contig of 1468 bp in length
* 98293 98392: gap of unknown length
* 98393 99454: contig of 1062 bp in length
* 99455 99554: gap of unknown length
* 99555 100912: contig of 1358 bp in length
* 100913 101012: gap of unknown length
* 101013 102195: contig of 1183 bp in length.

Query Match
Best Local Similarity 8.5%; Score 283.6; DB 2; Length 102195;
Matches 448; Conservative 0; Mismatches 274; Indels 0; Gaps 0;

QY 1769 agtaccatcaaggaagcactgtacaagtaccagtgccagtgaggccgcgcgaagaagt 1828
DB 29489 AGAAGAACAAGAAAGAAAGAAACAAGAAACAACAAGAAAGAAAGAAACAACAAGAAAGA 29548
QY 1829 gccaacggaagaattgatcgaactccaagaggacgatttcgaattagaaggaactcaga 1888
DB 29549 ACAAGAAGAAACAACAAGAAAGAAACAAGAAAGAAATCAAGAAGAAAGAAACAACAAGAAAGA 29608
QY 1889 agctccagaggaaggagaattagattagaagagagagagaccacgcgaaggaagacc 1948
DB 29609 AAAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 29668
QY 1949 aagagaaggagaccacagaagaagaagtgcagagaagaagaattagaggcaactccaga 2008
DB 29669 AAAAGAACAAGAAACAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 29728
QY 2009 ggacgatttcgaatagaagaaccaagagagagagagagagagaccggaagcgtagagggcga 2068
DB 29729 AGAAGAAGAAACAACAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 29788
QY 2069 agaaactgcagaaggagaagaagtgcgaagagggtaccctgcagaagaagtgcgaaga 2128
DB 29789 TGAAGAAGAAAGAAAGAGGAAAGGAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 29848
```

```
QY 2129 ggtacctgcagaagttagaagaagtggagaaggtaccagaagaagttagaagaggtaccgcc 2188
DB 29849 AGAAGAAGAAAGAAAGAAAGAAAGAAACAAGAAACAAGAAACAAGAAACAAGAAACAAGA 29908
QY 2189 agaagttagaagaagttagaagaggtaccagaagaagtggagaaggtaccagaagaagtgcga 2248
DB 29909 ACAACAAGGAAACAAGAAAGTGAACAAGAAACAAGAAAGAAACAAGAAACAAGAAACAAGA 29968
QY 2249 agaggttaccagaagaagttagaagaggtaccagaagaagtggagaaggtggaagaagtgcga 2308
DB 29969 AGAGGAACAGGAAACAAATGAAGATCAAGACGAAAGAACTAAGAAAGAAAGAGGAGGAGG 30028
QY 2309 agaagttagaggttaccagcggttagtagaagttagaaggtaccagcggttagtagaagaaggt 2368
DB 30029 AGGAGGAGGAGGAAACAAGAAACAAGAAACAAGAAACAAGAAACAAGAAACAAGAAACAAGA 30088
QY 2369 gccagaagaagttagaagaagaagaagaagaggaaggaacagtagtagagagaagaagtatt 2428
DB 30089 ACAACAAGAAAGAAACAAGAAAGAAACAAGAAACAAGAAACAAGAAAGAAAGAAAGAAAGA 30148
QY 2429 acaattagtaataccatcggaagaagatatataattagacaaacccaagaagacgaatt 2488
DB 30149 AGAAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGA 30208
QY 2489 ag 2490
DB 30209 AG 30210

RESULT 14
AC090533/C
LOCUS
DEFINITION
272545 bp DNA linear HTG 25-JAN-2002
Mus musculus clone RP23-33H13 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 45 unordered pieces.
AC090533
AC090533.4 GI:17976439
HTG: HTGS_PHASE1; HTGS_DRAFT.
VERSION
house mouse.
KEYWORDS
Mus musculus
SOURCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 272545)
Li, L., Montgomery, K.F., Grills, G., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
High Throughput Mouse Sequencing
AUTHORS
Unpublished
TITLE
2 (bases 1 to 272545)
Li, L., Montgomery, K.F., Grills, G., Chiu, D., Decker, J., Fusina, M.,
Goltz, J., Haider, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
Perera, A., Shim, C., Thomas, E. and Kucherlapati, R.
Direct Submission
JOURNAL
Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
Bronx, NY 10461, USA
COMMENT
On Dec 21, 2001 this sequence version replaced gi:13310871.
-----Genome Center
Center: Harvard Partners Genome Center
Web site:
http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous
e.html
Contact: gnkn@capecod.bwh.harvard.edu
-----Summary Statistics
Center project name: ADU
Sequencing vector: pUC18; L08752
Chemistry: Dye-terminator Big Dye; 100%
*Consensus quality: 257441 at least Q20
*Consensus quality: 253718 at least Q30
*Consensus quality: 247095 at least Q40
**Estimated insert size: agarose-pp - N/A
**Estimated insert size: 271665 - sum-of-contigs
Quality coverage: agarose-pp - N/A
```

Quality coverage: 6.8 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 30267: contig of 30267 bp in length
* 30268: contig of unknown length
* 30288 58400: contig of 28113 bp in length
* 58401 58420: gap of unknown length
* 58421 88864: contig of 30444 bp in length
* 88865 88884: gap of unknown length
* 88885 109889: contig of 21005 bp in length
* 109890 109909: gap of unknown length
* 109910 125133: contig of 15224 bp in length
* 125134 125153: gap of unknown length
* 125154 139549: contig of 14396 bp in length
* 139550 139569: gap of unknown length
* 139570 153410: contig of 13841 bp in length
* 153411 153430: gap of unknown length
* 153431 162567: contig of 9137 bp in length
* 162568 162587: gap of unknown length
* 162588 171044: contig of 8457 bp in length
* 171045 171064: gap of unknown length
* 171065 179304: contig of 8240 bp in length
* 179305 179324: gap of unknown length
* 179325 187059: contig of 7735 bp in length
* 187060 187079: gap of unknown length
* 187080 197598: contig of 10519 bp in length
* 197599 197618: gap of unknown length
* 197619 205966: contig of 8348 bp in length
* 205967 205986: gap of unknown length
* 205987 212449: contig of 6463 bp in length
* 212450 212469: gap of unknown length
* 212470 217752: contig of 5283 bp in length
* 217753 217772: gap of unknown length
* 217773 223943: contig of 6171 bp in length
* 223944 223963: gap of unknown length
* 223964 230362: contig of 6399 bp in length
* 230363 230382: gap of unknown length
* 230383 236251: contig of 5869 bp in length
* 236252 236271: gap of unknown length
* 236272 240572: contig of 4301 bp in length
* 240573 240592: gap of unknown length
* 240593 242789: contig of 2197 bp in length
* 242790 242809: gap of unknown length
* 242810 244695: contig of 1886 bp in length
* 244696 244715: gap of unknown length
* 244716 246969: contig of 2254 bp in length
* 246970 246989: gap of unknown length
* 246990 247870: contig of 881 bp in length
* 247871 247890: gap of unknown length
* 247891 248193: contig of 303 bp in length
* 248194 248213: gap of unknown length
* 248214 250122: contig of 1909 bp in length
* 250123 250142: gap of unknown length
* 250143 251367: contig of 1225 bp in length
* 251368 251387: gap of unknown length
* 251388 252749: contig of 1362 bp in length
* 252750 252769: gap of unknown length
* 252770 253640: contig of 871 bp in length
* 253641 253660: gap of unknown length
* 253661 255032: contig of 1372 bp in length
* 255033 255052: gap of unknown length
* 255053 257333: contig of 2281 bp in length
* 257334 257353: gap of unknown length
* 257354 259409: contig of 2056 bp in length
* 259410 259429: gap of unknown length
* 260459: contig of 1030 bp in length

* 260460 260479: gap of unknown length
* 260480 260829: contig of 350 bp in length
* 260830 260849: gap of unknown length
* 260850 261661: contig of 812 bp in length
* 261662 261681: gap of unknown length
* 261682 262532: contig of 851 bp in length
* 262533 262552: gap of unknown length
* 262553 263351: contig of 799 bp in length
* 263352 263371: gap of unknown length
* 263372 264217: contig of 846 bp in length
* 264218 264237: gap of unknown length
* 264238 265205: contig of 968 bp in length
* 265206 265225: gap of unknown length
* 265226 266270: contig of 1045 bp in length
* 266271 266290: gap of unknown length
* 266291 266749: contig of 459 bp in length
* 266750 266769: gap of unknown length
* 266770 267718: contig of 949 bp in length
* 267719 267738: gap of unknown length
* 267739 268576: contig of 838 bp in length
* 268577 268596: gap of unknown length
* 268597 270010: contig of 1414 bp in length
* 270011 270030: gap of unknown length
* 270031 271654: contig of 1624 bp in length
* 271655 271674: gap of unknown length
* 271675 272545: contig of 871 bp in length.
* 271675 Location/Qualifiers
1. .272545
/organism="Mus musculus"
/strain="C57BL6/J"
/db_xref="taxon:10090"
/clone="RP23-33H13"
/sex="male"
1. .30267
/note="assembly_name:Contig108"
30288. .58400
/note="assembly_name:Contig107"
clone_end:SP6
vector_side:left
58421. .88864
/note="assembly_name:Contig106"
88885. .109889
/note="assembly_name:Contig105"
109910. .125133
/note="assembly_name:Contig104"
clone_end:T7
vector_side:right
125154. .139549
/note="assembly_name:Contig103"
139570. .153410
/note="assembly_name:Contig102"
153431. .162567
/note="assembly_name:Contig101"
162588. .171044
/note="assembly_name:Contig100"
171065. .179304
/note="assembly_name:Contig99"
179325. .187059
/note="assembly_name:Contig98"
187080. .197598
/note="assembly_name:Contig97"
197619. .205966
/note="assembly_name:Contig96"
205987. .212449
/note="assembly_name:Contig95"
212470. .217752
/note="assembly_name:Contig94"
217773. .223943
/note="assembly_name:Contig93"
223964. .230362
/note="assembly_name:Contig92"
230383. .236251
/note="assembly_name:Contig91"

FEATURES
Source

Qy	2085	gaaagaagtgggaaggttacctgcagaagtagtaagaagtgggaagagggtacctgcagaagta	2144
Db	177896	GCAGGAGGACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	177837
Qy	2145	gaagaagtgggaaggttaccagaagaagttagaagaggttaccctcgacagaagttagaagaagt	2204
Db	177836	GGACGAGGACGAGCAGGAACAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCA	177777
Qy	2205	gaagagttaccagaagaagtgggaagaggttaccagaagaagtgggaagaggttaccagaaga	2264
Db	177776	GGAGGACGAGGAGCAGCAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCA	177717
Qy	2265	gtggaagaggttaccagaagaagtgggaagaggttagaagaagttagaagaagttagaagttacca	2324
Db	177716	GCAGCAGGAGGACGAGGAGGAGCAGGAACAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGCA	177657
Qy	2325	gcggttagtagaagttagaagttaccagcggtagtagaagaagaggtgccagaagaagttagaa	2384
Db	177656	GCAGGAGCAGGAGCAGCAGCAGGAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCAGGAGCACTGGGGCA	177559
Qy	2385	gaagaagaagaaggaagaaccagttagagaagaagaatgata	2426
Db	177596	GGACAGGAGCAGAAGGGCGGGGCGAGGAGTGGGAGTGGA	177555
RESULT	15		
AC018634/c			
LOCUS	AC018634	174241 bp	DNA linear PRI 28-JUL-2000
DEFINITION	Human Chromosome 7 clone RP11-243E12, complete sequence.		
ACCESSION	AC018634		
VERSION	AC018634.3	GI:9558607	
KEYWORDS	HTG.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 174241) Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and Olson,M.V. Large-scale Mapping and Sequencing of Human Chromosome 7 Unpublished 2 (bases 1 to 174241) Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M. Direct Submission Submitted (15-DEC-1999) Human Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA 3 (bases 1 to 174241) Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and Olson,M.V. Direct Submission Submitted (28-JUL-2000) Genome Center, University of Washington, Box 352145, Seattle, WA 98195, USA On Jul 28, 2000 this sequence version replaced gi:8050915. ----- Genome Center Center: University of Washington Genome Center Center Code: UWGC Web site: http://www.genome.washington.edu/UWGC/ Contact: uwghtgsu.washington.edu ----- Project Information Center project name: HsaChr7 Center clone name: RP11-243E12 (djsl56) ----- Summary Statistics Sequencing vector: M13; X02513; 100% of reads Chemistry: Dye-primer-amersham; 66% of reads Chemistry: Dye-terminator ET-amersham; 34% of reads Assembly program: Phrap; version 0.990319 Consensus quality: 173493 bases at least Q40 Consensus quality: 174139 bases at least Q30 Consensus quality: 174231 bases at least Q20 Insert size: 179271; 8.6% error; agarose-fp Insert size: 174241; sum-of-contigs Quality coverage: 6.32x in Q20 bases; agarose-fp Quality coverage: 6.50x in Q20 bases; sum-of-contigs		

 Overlapping Sequences:
 5': RP11-605P22 (UMGC:djs708)
 3': Mapping in progress

Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Double stranded (DS) coverage: 29%

DS or two chemistry coverage: 29%

Single stranded regions: 0.787% in 7 gaps

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

BgIII				EcoRI				NotI			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
2554	2514	8696	8906	12972	12987						
2067	2002	210	<800	1397	1379						
4648	4783	605	<800	4332	4323						
3573	3614	545	<800	6863	6887						
8078	8030	4828	4820	3199	3170						
4368	4326	774	782	1887	1838						
16815	16889	11779	11746	1024	1024						
2290	2300	2846	2851	4955	4937						
4064	4019	7395	7450	2501	2457						
12849	12884	15587	15482	167	<800						
3693	3850	4041	4019	4213	4186						
2740	2714	3827	3801	7678	7631						
3872	4019	3925	3916	3405	3413						
491	<800	2972	2950	2335	2385						
1890	1885	1716	1728	1344	1327						

6066	5966	18105	18158	1076	1043
867	823	4047	4019	467	<800
7550	7595	479	<800	938	991
1542	1531	7518	7450	3184	3170
5259	5070	115	<800	2431	2385
2061	2053	10695	10628	130	<800
3357	3364	1142	1121	2209	2189
7394	7324	4194	4160	2227	2231
813	823	3104	3123	1857	1838
5190	5070	1019	1012	8477	8468
308	<800	9558	9461	393	<800
4317	4326	976	973	1041	1043
2748	2714	1129	1121	139	<800
6726	6665	10185	10126	821	816
6831	6877	1487	1486	3610	3669
2355	2393	433	<800	618	<800
2172	2140	3371	3385	4153	4186
1916	2002	10484	10628	3167	3170
2396	2514	10143	10126	630	<800
2512	2514	8539	8577	2314	2231
4092	4326	663	<800	1846	1800
900	920	122	<800	3624	3669
1618	1614	174	<800	1026	1024
6696	6534	4551	4590	807	816
4460	4326	952	973	190	<800
5469	5309	7	<800	1808	1800
3387	3614			987	991
2195	2300			1806	1800
2555	2514			513	<800
1416	1402			4060	4186
610	<800			398	<800
225	<800			274	<800
932	920			983	991
1982	1885			722	726
29	<800			1638	1760
74	<800				
853	844				

```
----- 1183 1154
-----
987 1024
-----
70 <800
-----
575 <800

Query Match      8.3%; Score 278.4; DB 9; Length 174241;
Best Local Similarity 55.1%; Pred. No. 4.2e-31;
Matches 693; Conservative 0; Mismatches 551; Indels 14; Gaps 7;

QY 1231 gaacagaaacgtagtagggagaaacacagaagctgcagaaggagaagtatcag 1290
Db 146098 GAAAGAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 146039

QY 1291 aagaaactcagaagagagaagagtagtaggcaactccagaggtatttcgcattag 1350
Db 146038 AAGAAAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145979

QY 1351 atggaactacattagaaagaaacccgaagaaactgcagagaggaagaaacccgttagaggag 1410
Db 145978 ACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 145919

QY 1411 aagaaacgttagaggagaagaaacccgttagaggagaagagctgcagaaggagaagag 1470
Db 145918 GAGAAGGAGAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145859

QY 1471 agtttagggcaactccagagtagtgaacttccaattagaagaaccatcaggagaaggagaag 1530
Db 145858 GAGGAGACAGAGAGGAGGAGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 145799

QY 1531 gggaaaggagagagaggaggaggaaggaagcgtagtagtcagtcagtcagtcagtcagtcagtc 1590
Db 145798 AGGAAGGAGAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145739

QY 1591 ccgaacccgttagaagtagtagtgaactcctgcagcctgtcgaacccaatggtcgctcccaacgg 1650
Db 145738 AGGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145685

QY 1651 cagatgaacttatttcgttgcatttcattgaataacgatttaacgtatgcagcattacat 1710
Db 145684 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145625

QY 1711 cctttagccattattttaacaaatcctcaaggtatcctgcagtaggcagtaggcagtaggcagtag 1770
Db 145624 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145565

QY 1771 taccatcaaggagacacctgtacaagtaccagttgcagtaggcagtaggcagtaggcagtaggcag 1830
Db 145564 ATGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145505

QY 1831 caacggagaagtatgatcgaactccaagagacgatttcgaattagaaggaactgcagaag 1890
Db 145504 AGGAAGGAGAGAGTGAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145445

QY 1891 ctccagaggagagagaattagttattagaaggaaggaaggaacccaacggaagaagagccaa 1950
Db 145444 GAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145387

QY 1951 gagaagagagccaagagaggaagtgccagaagaagattagaggcaactccagagg 2010
Db 145386 AAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145328

QY 2011 acgatttcctatagaagaacccaagagagaagtagagaagaaacccgttagagggcgaag 2070
Db 145327 AAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145268

QY 2071 aaactgcagaagagagaagtagtgaagaggtacctgcagaagtagagaagtaggaagagg 2130
Db 145267 AGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145208
```

```
QY 2131 tactgcagaagtagaagaagtgtgaagaggtaccagaagaagtagaa--gaggtaccgcg 2188
Db 145207 AAGAGGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145148

QY 2189 agaagtagaagaagtagtgaagaggtaccagaagaagtggaagaggtaccagaagaagtgg 2248
Db 145147 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 145088

QY 2249 agaggtaccagaagaagtggaagaggtaccagaagaagtggaagaggtggaagaagtag 2308
Db 145087 CGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145030

QY 2309 agaagtagaggtaccagcggttagtagaagtagaagtagaccag-cggttagtagaagaagagg 2367
Db 145029 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 144970

QY 2368 tgcagaagaagtagaagaagaagaagaaggaaggaaggaaggaaggaaggaaggaagtag 2427
Db 144969 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 144910

QY 2428 tacaattagtaataccatcggaagaagatacataattagacaacccaagaaagacga 2485
Db 144909 AAGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 144852

RESULT 16
AL645746
LOCUS
DEFINITION
AL645746
ACCESSION
VERSION
AL645746.2 GI:17148444
HTG; HTGS_PHASE1.
KEYWORDS
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
Sims,S.
REFERENCE
1 (sites)
AUTHORS
DIRECT SUBMISSION
TITLE
Submitted (11-JAN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Nov 28, 2001 this sequence version replaced gi:17017892.
COMMENT
----- Genome Center
Center: UK Medical Research Council
Center code: UK-MRC
Web site: http://mrcseq.har.mrc.ac.uk
Contact: mouse@har.mrc.ac.uk
----- Project Information
Center project name: bM153B6
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Consensus quality: 279809 bases at least Q40
Consensus quality: 280516 bases at least Q30
Consensus quality: 281062 bases at least Q20
Insert size: 281711; sum-of-contigs
Quality coverage: 12.39x in Q20 bases; sum-of-contigs Quality
coverage: 18.52x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
FEATURES
Location/Qualifiers
1..282611
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="13"
/clone="RP23-153B6"
```


Db 1870 AGTTAAAAAAGTGTAAAAAAGTAAAAAAGGAAATTAAGCAAAATGATCTGAAGG 1929
Qy 548 tggcgatatgaagcaggggaagaactggtgatttgaagcaggggaagaactggcga 607
Db 1930 TAACGATAAAGTGAAGGACCGACAGATAATTAATTAAGAGAGTAAAGAGCAATTAATA 1989
Qy 608 ttggaagcaggggaagaactggcgatttgggaagcaggggaagaactggtgatttggga 667
Db 1990 ACAAGTAGAAGATGGAATTAAGAAATGATCTGAAGGTAAACGATAAAGTGAAGGGCC 2049
Qy 668 agcaggggaagaactggcgatttgggaagcaggggaagaactggagatcggaactga 727
Db 2050 AGAAATTAATCTGAAGAGTAAAGAGAAATTAAGAAATTAAGAAATTAAGAAATTAAG 2109
Qy 728 agaagggcgaactggcgatcggaactggaactgg-----agcaactgtgataga 781
Db 2110 AGAAATTAATCTGAAGTAAAGTAAAGTGAAGGCGCGCAAAATTAATACTGAAGAGT 2169
Qy 782 cacagaagatagttcagctgtaggagcagaaaaagtacattgttcttgcctcaagaactgt 841
Db 2170 AAAAGAGAAATTAAGAAATCTAGAAAGAGGAATTAAGAAATTAAGAAATTAAGAA 2229
Qy 842 acaactgcgagatagatgacccctcttgggaagtatttggataaaagataataat 901
Db 2230 AGATAAAGTGTAGGACAGAAATAATACTGAAGAGTAAAGAGGAATTAAGAGAA 2289
Qy 902 tgatcatattaaagatttcgagccacttccgaacaaattgtgcggttactgctaaaca 961
Db 2290 TGATACTGAAATTAAGATAAAGTATAGCAGCAGAAATGATACTGAAGAGTAAAAA 2349
Qy 962 tgttacgggacaagaattgccaatgaacctgtaccattaccgtggcagagagcccg 1021
Db 2350 AGAAATTTGAAAAACAGAGAAAGAAATAAGAAATTAATTTGAAATTAAGAGATAT 2409
Qy 1022 gcaagtaccagcgaagaattagatgccactccagagatgacttgcattagatgttac 1081
Db 2410 AGTAATTTGGACAGAGTAAATATAGAAAGTAAAGAAATTAATTAAGAAAGTAGA 2469
Qy 1082 agaattcccgaggaagtagaattagtagtagagagcgaactgaagaagaactaac 1141
Db 2470 AAAAGGAATTAAGAAATCATACTGAAAGTAAAGATAAAGTATAGGACAGAAATAAT 2529
Qy 1142 ggaagtgggaccacggaagaagaccacccgaagaattagatgccactccagagatgg 1201
Db 2530 AGTTGAAGAGTAAAGAGAAATTTGAAACCAAGTAGAAGAGGAATTAAGAGAAATGA 2589
Qy 1202 atttcgcatlagcgaactgcgaagagagaacagagaacgtagagggagagaac 1261
Db 2590 TACTGAAGTAAAGATAAAGTGTAGCACAAAGTATGAAGGAGATGTTAATGAAGA 2649
Qy 1262 agaagaagctgcgaaggagaagtatcagaagaactccagaaggaggaaggaagaac 1321
Db 2650 AGGTCCGAAAAAAGATTAAGTACAAAAACAGGAAAAAGTAAAGAACTTAAGAAAA 2709
Qy 1322 gcaactccagagatgatttcgattagatgaactacattagaagaaccggaagaac 1381
Db 2710 AGTAAAAAAGAGTTAAAAAAGAGTAAA-----AAAAAGAAATTAATGAAGTGAAGAAA 2766
Qy 1382 tgcagaaggagaagaacccttagaggagaagaacccttagaggagagaagaacccttaga 1441
Db 2767 AGATAATTTTAGGAAAGAAATTAATGAAGAGAGTGTAAATGAAGAAATACCCGAAA 2826
Qy 1442 ggaagaagaagctgcagaaggagaagagtaggttagggcgaactccagagatgacttcca 1501
Db 2827 CAAGATTAAGAGATAGAACAGAAAAAGAAAAAGAGAGTAAAGAGAAAGTAAAGAA 2886
Qy 1502 attagaagaaccatcagagagaaggaggaaggagagaggaaggaggaaggaggaagg 1561
Db 2887 TAAAGAAAAAGAGAGTAAAGAAAAAGAGAGTAAAGAAAAAGAGAGAGTAAAGAA 2946
Qy 1562 agaagcgttagtagcagtagtggccgaacccggttagagtagtactcctgtcca 1621
Db 2947 AAAAGAGAGTAAAGAAAAAGAGAGTAAAGAAAAAGAGAGTAAAGAAAAAGAA 3006

Qy 1622 gcgtgtcaaccaattggtcgctcccaagcagatgaacttttattgtgtatctttaga 1681
Db 3007 AGAAGTAAAAAAGATAACCGAAAGCAAAAGATAAAGAGATAGAACAAAGAAAAAGAAA 3066
Qy 1682 taacgatttaacgtatgcagacattacatcttcttgagccattttaaacaatctctcaa 1741
Db 3067 AGAAGAGTAAAAAAGAGTTAAAGAAAAAGATACCGAAAAAAGAAAAAGTAAAGTAGAGACA 3126
Qy 1742 ggaactgtatgcagagagggctgttaacagtaccatcaagaaggaagcaccctgtacaagtacc 1801
Db 3127 AGAAATTAATTAAGAGAAATAAAAAAAGAGTTAAAAAAGAGTAAAAAAGAAATAA 3186
Qy 1802 agtggcagtagggcccgcgcaagaagtgcacaagcgaagaattgtgccaactcccaagagga 1861
Db 3187 TAAAAATGAACCAAGATAATGTGTAGTACAGAAATAATGAACGAGATGTTACGAAA 3246
Qy 1862 cgatttcgaattagaaggaaactccagagctcagaggggaaggaattagattagaagg 1921
Db 3247 AGATACCGCAAAACAAAGATTAAGGTGATAGAACAAAGAAAAAGAAAAAGAAAGTAAAAA 3306
Qy 1922 agaagagaacccaagaaagagagccaagagagagcgaagagaggaaggaagtgc 1981
Db 3307 AAAAGAGAGTAAAAAGAAAAAGAAAGTAAAGAAAAAGAAAGTAAAGAAAAAGAA 3366
Qy 1982 agaagaagattagagggcaactccagaggcgatttcgaattagaagaacccaacagagga 2041
Db 3367 AGAAGTAAAAAAGAAAAAGATACCGAAAGCAAAAGATAAAGAGATAGAACAAAGAAAAAGAAA 3426
Qy 2042 agaagttagaagaacccgttagaggcggaagaactcgaagaaggagaagaactggaagaggt 2101
Db 3427 AGAAGAGTAAAAAGAGTTAAAGAAAAAGATACCGAAAAAAGAAAAAGTAAAGTAGAGACA 3486
Qy 2102 acctgcagaagttagaagaagtgggaaggttacctgcagaagttagaagaagtgggaagaggt 2161
Db 3487 AGAAATTAATTAAGAGAAATAAAAAAGAGTTAAAAAAGAGTAAAAAAGAAATAA 3546
Qy 2162 accagaagaagttagaagggttaccgcagagtagaagaagtgggaagaggtaccagaaga 2221
Db 3547 TAAAAATGAACCAAGATAATGTGTAGTACAGAAATAATGAACGAAAGTGTATACGA 3606
Qy 2222 agtgggaaggttaccagaagaagtgaagaggtaccagaagaagtgggaagatccaga 2281
Db 3607 AAAAGTATCCCGCAAAACAAAGTAAAGTGTAGAACAAAGAAAAAGAAAAAGAAAGTAAA 3666
Qy 2282 agaagtgggaagaagtgggaagaagttagaagaagttagaggttaccagcggtagtagaagtaga 2341
Db 3667 AGAAAAAGAGAGTAAAGAAAAAGAAAGTAAAGAAAAAGAAAGTAAAGAAAAAGAAA 3726
Qy 2342 agtaccagcggttagttagaagaaggtgcagagaagtagaagaaggaagaag----- 2394
Db 3727 AGATACCGAAAGCAAGATTAAGAGATAGAACAAAGAAAAAGAAAAAGAAAGTAAAAAGA 3786
Qy 2395 -----aagaggaagaaccagtagagaagaaggttattacaattagtaaccatcgga 2449
Db 3787 AGTTAAAGAAAAAGATACCGAAAAACAAAGATAAAGTGTAGGACAGAAATAATATATAGA 3846
Qy 2450 agaagatatcaattagacaacccaagaagagcgaattaggctctggaattttattctat 2509
Db 3847 AGAAATAAAAAAGAGTTAAAAAAGAGTAAAAAAGAAAAATAATAAAAAAGAAACAA 3906
Qy 2510 catcgacatgcactccaagacgttccaaaggaa---tttatggagaagaagaagaac 2566
Db 3907 AGATAATGTGTAGTACAGAAAAATTAATGAACGAAGTGTAAACGAAAAAGATACCCGAAA 3966
Qy 2567 tgcagtgattcattgaaaccagaagatttgcagaaggaagattccacaattctacagaagt 2626
Db 3967 CAAGATTAAGGTATAGAACAAAGAAAAAGAAAGAGAGTAAAGAAAGTAAAGAAAA 4026
Qy 2627 gctcacattcatccaagggcctagaagcgactgggaacgattagaagtagagcttaataa 2686
Db 4027 AGATACCGTAAACAAAGATTAAGTGTAGGACAAAGAGTAAATAAGAAAGTAAAAAGA 4086


```
* 51200 53719: contig of 2520 bp in length
* 53720 53819: gap of 100 bp
* 53820 56012: contig of 2193 bp in length
* 56013 56112: gap of 100 bp
* 56113 57226: contig of 1114 bp in length
* 57227 57326: gap of 100 bp
* 57327 77403: contig of 20077 bp in length
* 77404 77503: gap of 100 bp
* 77504 79994: contig of 2491 bp in length
* 79995 80094: gap of 100 bp
* 80095 82923: contig of 2829 bp in length
* 82924 83023: gap of 100 bp
* 83024 86207: contig of 3184 bp in length
* 86208 86307: gap of 100 bp
* 86308 90457: contig of 4150 bp in length
* 90458 90557: gap of 100 bp
* 90558 96908: contig of 6351 bp in length
* 96909 97008: gap of 100 bp
* 97009 102861: contig of 5853 bp in length
* 102862 102961: gap of 100 bp
* 102962 107881: contig of 4920 bp in length
* 107882 107981: gap of 100 bp
* 107982 113543: contig of 5562 bp in length
* 113544 113643: gap of 100 bp
* 113644 118163: contig of 4520 bp in length
* 118164 118263: gap of 100 bp
* 118264 135271: contig of 17008 bp in length
* 135272 135371: gap of 100 bp
* 135372 151730: contig of 16359 bp in length
* 151731 151830: gap of 100 bp
* 151831 171146: contig of 19316 bp in length
* 171147 171246: gap of 100 bp
* 171247 174019: contig of 2773 bp in length.
```

FEATURES

source

```
1. .174019
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="15"
```

```
/map="15"
```

```
/clone="RP11-228G18"
```

```
/clone_lib="RPC1-11 Human Male BAC"
```

```
1. .13487
```

```
/note="assembly_fragment"
```

```
clone_end:SP6
```

```
vector_side:left"
```

```
13588..14626
```

```
/note="assembly_fragment"
```

```
14727..15793
```

```
/note="assembly_fragment"
```

```
15894..17057
```

```
/note="assembly_fragment"
```

```
17158..18291
```

```
/note="assembly_fragment"
```

```
18392..20046
```

```
/note="assembly_fragment"
```

```
20147..21386
```

```
/note="assembly_fragment"
```

```
21487..22760
```

```
/note="assembly_fragment"
```

```
22861..23953
```

```
/note="assembly_fragment"
```

```
24054..25081
```

```
/note="assembly_fragment"
```

```
25182..26357
```

```
/note="assembly_fragment"
```

```
26458..27574
```

```
/note="assembly_fragment"
```

```
27675..28756
```

```
/note="assembly_fragment"
```

```
28857..30554
```

```
/note="assembly_fragment"
```

```
30655..32124
```

```
/note="assembly_fragment"
```

```
misc_feature 32225..33634
```

```
misc_feature /note="assembly_fragment"
```

```
33735..35520
```

```
/note="assembly_fragment"
```

```
Query Match 7.7%; Score 255.8; DB 2; Length 174019;
```

```
Best Local Similarity 48.6%; Pred. No. 7.6e-28;
```

```
Matches 918; Conservative 0; Mismatches 942; Indels 27; Gaps 7;
```

```
QY 531 gaagagaagaagaactgcgcataatgaagcagcaggaggaagaactggtgattgggaagca 590
```

```
Db 66198 GGAGGAGATGTGGAGCAGGAGGAGGAGCTACGGGACGACGAAAGAAAGATACGGAAGCA 66257
```

```
QY 591 ggggaagaactggcgtatttgaagcagggaagaactggcgatttgggaagcaggagaa 650
```

```
Db 66258 GGAGGAGAA-----GATGTGGAGACAGGAGGAGAGACTCGG---GAGCAGGAGGGGAA 66308
```

```
QY 651 gaagctgggtgatttgaagcagggggaagaactggcgatttgggaagcaggaggaagaact 710
```

```
Db 66309 GATGCGGGACGACGAGGAGGAGAAAGATGCGGACACAGGAGAGAGGCTCGGAGAGGAGAA 66368
```

```
QY 711 ggagatgcggaactgaagaagcagcaactggagatgcggaactggaactggaactgagcaact 770
```

```
Db 66369 GGAAGCTCGGGACGACGAGGAGGAGGAGCTGCGGGAGCAGAAAGCTCGGGAGGAGGAGGA 66428
```

```
QY 771 gttatgtagacacagaagatagttcagctgattggagcagaaaaagatactgtctctgct 830
```

```
Db 66429 GCAGATGCAGGAGCAGGAGGAGGAGAAAGATGTGGAGCAGGAGGAGGAGGAGGAGCA 66488
```

```
QY 831 caagaataatgacaactgccgatagtaataatgacccctcttggaaagtatttggataaa 890
```

```
Db 66489 AGAGAGATGTGGAGCAGGAGGAGGAGGAGGCTGTGGGACGAGGAGGAGGAGGAGGAG 66548
```

```
QY 891 gataataattttgatcatattaaagatttcgagcactattcgaaacaaatttggcgggt 950
```

```
Db 66549 GGAGCAGAGAGATCGGGACCAACAGGAGGAGGATGTGGGACGAGGAGGAGGAGGAGGAG 66608
```

```
QY 951 actgctaaacatttccgggacaagaattgccaatgaaacc---tgtaccattaccagt 1007
```

```
Db 66609 GAAGGAGGAGAGGATCGGGAGCAGGAGGAGGAGAGATGTGGGACGAGGAGGAGGAGGAG 66668
```

```
QY 1008 gcagaagagcccgccgaagtaccagcgggaagaattagatccactccagaggatgacttc 1067
```

```
Db 66669 GGAGAGAGAGATCGAGGAGCAGGAGGAGGAGAGACGCGGGACGAGGAGGAGGAGGAGGAG 66728
```

```
QY 1068 gcattagattttacagaatctcccgaggaa---gtagaattagttatttagtaagaggca 1124
```

```
Db 66729 GGAGGAGAGGATACGGGAGCGGGAGGAGAGAGATCGGGGAAAGAGAGGAGGAGGAGGAG 66788
```

```
QY 1125 actgaagaagaatcaacggaagtgggaccaacggaagaagaccacccggaagattagat 1184
```

```
Db 66789 GCAGGAGGAGAGATGCAGAAAGCAGGAGGAGGAGAAATATGTGGGAGCAGGAGGAGGAG 66848
```

```
QY 1185 gccactccagaggatgatttcgatttagcgaactgcagaagaggaagacacagaagaac 1244
```

```
Db 66849 GCAGCAGCAGAGGCTCCCGAACAGAGAGGAGAGAGCTGTGGGAAACAGGAGGAGGAGGAG 66908
```

```
QY 1245 gtagagggaagaacacagaagaagctgcagaaggagaagatatacagaagaactcccgaa 1304
```

```
Db 66909 GCAGGAGGAGAGATATGGGAGCAGGAGGAGGAGAGATACACGGGACCGAGGAGGAGGAG 66968
```

```
QY 1305 ggagaagaagtttagagcgaactccagaggatgatttcgca-ttagatggaactacatt 1363
```

```
Db 66969 GCAGGAGAGAAAGATGTGGCGGCGGAGGAGAGATGCGGGAGCAGGAGGAGGAGGAGGAG 67028
```

```
QY 1364 agaagaacccaagaactcagaaggagagaacacccgtagagggaagaagaacccgtaga 1423
```

```
Db 67029 AGGAGAGGCTGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGATC 67088
```

```
QY 1424 gggagaagaacccgt---agagggaagaagctgcagaaggagaagaagatttagagggc 1480
```

```
Db 67089 AGGAGCAGAGAGATGTGGGACCGAGGAGGAGGATGTGGGAGCAGGAGGAGGAGGAGGAGG 67148
```

QY	1481	aactccagaggaagtagcttccaattagaagaaccatcaggagaaggagaaggggaaagaga	1540
Db	67149	AGAAAGGAGGAGGAAGTACCGGAGCAGAGAAGATGTGGCAGCAGGTGCGAAGATCGGG	67208
QY	1541	aggagaagggaaggaagaaacgcttagtagcagtgccagtagtgccgaacccggt	1600
Db	67209	AGAGAAGAAGACGCGAGGACGAGGAAGAAAGATGTGGACCCAGGAGGAGAAGATCGGAG	67268
QY	1601	agaagtagtgactcctctgctcagcctgtcaaacccaatggctgcgtctcccaacgagatgaac	1660
Db	67269	AGGAGGAGGAGATCGGGAGCGGGAAGAAGATGCGGGAGGAGGAGGAGATCGCGG	67328
QY	1661	tttatcttgtagatcttagataacagatttaacgtagtcgaacattacatccctttgagcc	1720
Db	67329	AGCAGGAGGAGAAGATGCGAGGACGAGGAAGAAAGATGCGGGAGCAGGAGGAGAAGATGT	67388
QY	1721	attatttaacaaatcctcaagatcctgatcagagagaggtgttaacagtaccatcaa	1780
Db	67389	GGAGGAGGAAGAAGATGCGAGGACGAGGAGGAAGATGTGGGAGCAGGAGGAAGA	67448
QY	1781	ggaagcacctgtacaagtacagtgccagtagggcccgccgaagaagtgccaaacggaaga	1840
Db	67449	TGTGGGAGCAGGAAGATGTGGGAGCAGCAGAGCGTGTCCGGAACAGAGGAGAGC	67508
QY	1841	attgatgcaactccaagagagc-----atttgaattagaagaactgcagaagct	1892
Db	67509	TGTGGGAACACGAGAAGATGCAGGACGAGGAGATCAGGAGCAGAAGATGTGGGACCGAA	67568
QY	1893	ccagagaaggaagaattagtaggaaggaaggaagaaaccaacggaagaagaccaga	1952
Db	67569	GGAGAGGATGTGGGACGAGCAGCAGAGGTGTCCGGGAAGAGGAGGAGAAGATGCGGGAGCA	67628
QY	1953	gaaggagagccaagaagaagtgccagagaagaataagaggcaactccagagagac	2012
Db	67629	GAAGATGTGCACAGGTGGGAAGATGCGGGAGAGAAGAACACCGCAGGAGGAGAAT	67688
QY	2013	gatttcgaattagaagaaccaacagagagaagaagtagaagaacccgtagaggggcgaaga	2072
Db	67689	GCAGGAGCAGGAAGAAGATGCAGGAGCAGGAGGAGATGATGTGGGAGCAGGAAGAA	67748
QY	2073	actgcagaaggaagaagtggaagaggtaccagaagaagtagaagaaggtaccgcgagaa	2132
Db	67749	GATTCAGGACGAGGAGGAAGATGTGGGAGCAGGAGGAGGAAGATGCGGGAGCAGGAGGA	67808
QY	2133	cctgcagaagtagaagaagtgggaagaggtaccagaagaagtagaagaaggtaccgcgagaa	2192
Db	67809	GAAGATGCGGAGCAGGAGGAGAAGATGCAGGCGCAGGAGGAGCAAGATGCGGGAGCAGGA	67868
QY	2193	gtagaagaagtgggaagaggtaccagaagaagtggaagaggtaccagaagaagtggaagag	2252
Db	67869	GGAGAAGATGCGGGCCCATGAGGAGAAGATGCGGGAGCAGGAGGAGAAGATGCGGGGCCA	67928
QY	2253	gtaccagaagaagtgaagaaggtaccagaagaagtggaagaagtggaagaagtagaaga	2312
Db	67929	GGAGGAGAAGATGTGGGCCAGGAGGAGAAGATGTGGGGCAGGAGGAGAAGATGTGCGGG	67988
QY	2313	gtagaggtaccagcggttagtagaagtagaagtagaccagcggttagtagaagaaggtgcca	2372
Db	67989	GCAGGAGGAGATCGAGAGAAGGAGGAGGAGATACGAGATCAACAAGAGAAGATGCGAGGA	68048
QY	2373	gaagaagtagaagaagaagaagagag	2399
Db	68049	GAGGCTGCCAGACGAGGAGGCGGTG	68075

RESULT 22
AC106251
LOCUS
DEFINITION Rattus norvegicus clone CH230-201A10, *** SEQUENCING IN PROGRESS
AC106251
VERSION AC106251
KEYWORDS HTG; HTGS_PHASE1.

SOURCE ORGANISM

Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

REFERENCE AUTHORS

1 (bases 1 to 170276)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brivea,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F.,
Carter,M., Cavazos,S.R., Checko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Duggan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsif.F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,L., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegh,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,M., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Ogih,M., Okuwonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoshkar,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalob,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE JOURNAL

2 (bases 1 to 170276)
Unpublished
Worley,K.C.
Direct Submission
Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKHO
Center clone name: CH230-201A10
----- Summary Statistics
findPhrapList
Assembly program: Phrap; version 0.990329First call to
Consensus quality: 139818 bases at least Q40
Consensus quality: 147078 bases at least Q30
Consensus quality: 153529 bases at least Q20
Estimated insert size: 145171; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.2x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 13297)
AUTHORS	Voehringer,D., Kaufmann,M. and Pircher,H.
TITLE	Genomic structure, alternative splicing, and physical mapping of the killer cell lectin-like receptor G1 gene (KLRG1), the mouse homologue of MAFA
JOURNAL	Immunogenetics 52 (3-4), 206-211 (2001)
MEDLINE	2115136
PUBMED	11220622
REFERENCE	2 (bases 1 to 13297)
AUTHORS	Voehringer,D., Kaufmann,M. and Pircher,H.
TITLE	Direct Submission
JOURNAL	Submitted (31-OCT-2000) Immunology, Institute for Medical Microbiology and Hygiene, Hermann-Herderstr. 11, Freiburg 79104, Germany
FEATURES	Location/Qualifiers
source	1..13297
	/organism="Mus musculus"
	/strain="129/SvevTACfBr"
	/db_xref="taxon:10090"
	/sex="female"
	/tissue_type="spleen"
mRNA	join(<973..1054,4040..4144,5499..5668,10141..10238, 12303..>12414)
	/gene="KlrG1"
gene	/product="killer cell lectin-like receptor G1"
	<973..>12414
	/gene="KlrG1"
CDS	join(973..1054,4040..4144,5499..5668,10141..10238, 12303..12414)
	/gene="KlrG1"
	/note="MAFA/KLRG1"
	/codon_start=1
	/product="killer cell lectin-like receptor G1"
	/protein_id="AAK40082.1"
	/db_xref="GI:13810896"
REFERENCE	/translation="MADSSIYSTLELPEAPQVQDESFWKLAVLHRLPILSRFAMVALG LFLVLSLMYQRIVCCSKSDTSCHCPSLPILTRNGSHGYLFMSKKDWNSSLKF CADKGSHLLITFDNQKILTFGLVDGFYIGLIRNDIDGWREWGCPALSLRILTNSLIQ RCGAIHRNGLQASCSCEVALQWICKVLVY" BASE COUNT 3447 a 2838 c 2886 g 4108 t 18 others
ORIGIN	
	Query Match 7.6%; Score 252; DB 10; Length 13297;
	Best Local Similarity 64.4%; Pred. No. 3e-27;
	Matches 375; Conservative 0; Mismatches 207; Indels 0; Gaps 0;
QY	1835 ggaagaattgatgcaactccaaagaggcagtatttcgaatttagaaggaaactgcagaagctcc 1894
Db	11314 GGGGGAGGGAGGAGAGCGGGGAAGCAAGAAGCGCAGGCGGAGCGGAGGAGGAGGA 11255
QY	1895 agaggaaggagaattagtattagaaggagaagagaccacacgaagaagagccaagaca 1954
Db	11254 GGAGTCGGGGACAAGAACGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 11195
QY	1955 agagagccaacagagaaggaaagtgccacagaagaataatagggcaactccagaggcacg 2014
Db	11194 AGAAGAAGAAGAACGAAGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 11135
QY	2015 tttoaataggagaaccacagagagaagaagttagaagaacogtagaggcgcaagaac 2074
Db	11134 AGAAGAAGAAGAAGAAGAACGCAAGCAAGCAAGCAAGCGGCCGCAAGAAGAAGAAGA 11075
QY	2075 tcagaaggagaagaagtggaaagggtacctgcagaagttagaagaagtggaaagggtacc 2134
Db	11074 AGAAGAAGAAGAAGAAGCAAGCAAGCAAGCAAGCAAGCGGCCGCAAGAAGAAGAAGA 11015
QY	2135 tcagaagttagaagaagtggaaagggttaccagaagaagttagaagaagtgcacgaagt 2194
Db	11014 AGAAGAAGAAGAAGAAGAACGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGAAGA 10955
QY	2195 agagaagtggaaagggttaccagaagaagtggaaagggttaccagaagaagtggaaagg 2254

1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcoRI				HindIII				BglII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----
8696	8850	399	<800	2304	2280						
-----	-----	-----	-----	-----	-----						
6	<800	6382	6347	2067	2069						
-----	-----	-----	-----	-----	-----						
1907	1895	512	<800	5518	5637						
-----	-----	-----	-----	-----	-----						
1854	1895	449	<800	6918	7002						
-----	-----	-----	-----	-----	-----						
2849	2922	2322	2276	2029	2069						
-----	-----	-----	-----	-----	-----						
3722	3806	5951	5995	2378	2425						
-----	-----	-----	-----	-----	-----						
658	<800	7224	7291	2089	2069						
-----	-----	-----	-----	-----	-----						
26398	26593	1005	1033	550	<800						
-----	-----	-----	-----	-----	-----						
537	<800	1020	1033	4191	4187						
-----	-----	-----	-----	-----	-----						
1542	1483	3445	3448	3839	3773						
-----	-----	-----	-----	-----	-----						
1103	1096	2131	2126	1647	1646						
-----	-----	-----	-----	-----	-----						
1928	2045	2237	2276	629	<800						
-----	-----	-----	-----	-----	-----						
14029	13827	6625	6652	1932	1918						
-----	-----	-----	-----	-----	-----						
6707	6816	6349	6347	18266	18041						
-----	-----	-----	-----	-----	-----						
1884	1895	6669	6652	3099	3165						
-----	-----	-----	-----	-----	-----						
6048	6029	1943	1916	2900	2928						
-----	-----	-----	-----	-----	-----						
2150	2209	6376	6347	111	<800						
-----	-----	-----	-----	-----	-----						
13944	13827	5398	5641	2425	2425						
-----	-----	-----	-----	-----	-----						
4275	4287	599	<800	3767	3773						
-----	-----	-----	-----	-----	-----						
2057	2209	4055	4127	1506	1485						
-----	-----	-----	-----	-----	-----						
2293	2209	299	<800	3639	3773						
-----	-----	-----	-----	-----	-----						
6448	6393	20065	20079	5635	5637						

FEATURES

source
Location/Qualifiers
1. .179149
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-198E23"
/clone_lib="RPC1 human BAC library 11"
complement(3542..3629)
/standard_name="SWSS3607"
complement(87234..87324)
/standard_name="SWSS3670"
115832..116064
/note="Single subclone region"
130498..130537
/note="Single subclone region"

STS
STS
misc_feature
misc_feature

-----	9916	-----	6964	6903	-----	6894	7002
-----	7673	-----	5766	5641	-----	11819	11807
-----	115	-----	4281	4127	-----	837	821
-----	1490	-----	7844	7901	-----	2058	2069
-----	1071	-----	1032	1033	-----	3280	3312
-----	18341	-----	1591	1546	-----	3780	3773
-----	7618	-----	955	954	-----	623	<800
-----	1733	-----	330	<800	-----	202	<800
-----	1478	-----	1927	1916	-----	2950	2928
-----	7616	-----	6261	6347	-----	10775	10621
-----	561	-----	445	<800	-----	12114	11807
-----	19080	-----	3869	3877	-----	9445	9461
-----	114	-----	4148	4127	-----	1694	1646
-----	7	-----	3828	3877	-----	3745	3773
-----		-----	4001	4127	-----	674	<800
-----		-----	1561	1546	-----	9893	9817
-----		-----	3188	3192	-----	3225	3165
-----		-----	4960	4943	-----	998	1011
-----		-----	9109	8931	-----	1318	1283
-----		-----	929	954	-----	94	<800
-----		-----	7844	7901	-----	796	821
-----		-----	535	<800	-----	111	<800
-----		-----	5397	5408	-----	1029	1011
-----		-----	1905	1916	-----	1491	1485
-----		-----	113	<800	-----	20564	20918
-----		-----	5596	5641	-----		
-----		-----	750	735	-----		
-----		-----	1264	1248	-----		

ambiguous, there is an annotation using the 'unsure' feature key. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6> RP11-328C17 is from the library RPCI-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACe3.6

IMPORTANT: This sequence is not the entire insert of clone RP11-328C17. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-328C17 is at 132449 in this sequence. The true left end of clone RP11-233K4 is at 64893 in this sequence. The true right end of clone RP3-416J7 is at 100 in this sequence. The true right end of clone RPI-125A24 is at 61243 in this sequence.

FEATURES

Location/Qualifiers

Source

1..132449
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP11-328C17"
/clone_lib="RPCI-11.2"

repeat_region
1..1562
/note="LIM4 repeat: matches 4133..5719 of consensus"
repeat_region
1563..1641
/note="MLT1A2 repeat: matches 1..80 of consensus"
repeat_region
1669..2199
/note="LIM4 repeat: matches 3396..3937 of consensus"
repeat_region
2200..2474
/note="LIM4 repeat: matches 38..312 of consensus"
repeat_region
2475..4279
/note="LIM4 repeat: matches 1686..3396 of consensus"
repeat_region
4306..4589
/note="LIM4 repeat: matches 1254..1544 of consensus"
misc_feature
complement(4704..5324)
/note="match: GSS: Em:AQ628183"
repeat_region
4799..4938
/note="LIM4c repeat: matches 785..924 of consensus"
misc_feature
complement(4809..5326)
/note="match: GSS: Em:AQ774852"
misc_feature
complement(4860..5272)
/note="match: GSS: Em:AQ787689"
repeat_region
4943..5094
/note="LIM4c repeat: matches 272..420 of consensus"
misc_feature
complement(5020..5329)
/note="match: GSS: Em:AQ494601"
repeat_region
5142..5435
/note="MLT1D repeat: matches 170..428 of consensus"
repeat_region
5436..5744
/note="AluSx repeat: matches 1..310 of consensus"
repeat_region
5745..5869
/note="MLT1D repeat: matches 428..568 of consensus"
repeat_region
6450..6686
/note="MIR repeat: matches 7..261 of consensus"
repeat_region
6916..7994
/note="LIM4 repeat: matches 3582..4688 of consensus"
repeat_region
7984..8050
/note="LIMB8 repeat: matches 5808..5884 of consensus"
repeat_region
8064..8169
/note="LIM4 repeat: matches 3479..3593 of consensus"
repeat_region
8178..8224
/note="LIMB4 repeat: matches 6124..6169 of consensus"
repeat_region
8225..8525
/note="AluSx repeat: matches 1..301 of consensus"
repeat_region
8526..8915
/note="LIMB4 repeat: matches 5725..6124 of consensus"

misc_feature
complement(8862..9353)
/note="match: GSS: Em:AQ355497"
repeat_region
9029..9150
/note="2 copies 61 mer 95% conserved"
misc_feature
complement(9052..9484)
/note="match: GSS: Em:AQ175845"
misc_feature
9388..9675
/note="match: GSS: Em:AQ135206"
misc_feature
9487..9840
/note="match: GSS: Em:AQ171684"
misc_feature
9491..9932
/note="match: GSS: Em:AQ147759"
repeat_region
9524..9697
/note="MER53 repeat: matches 1..189 of consensus"
repeat_region
10046..10143
/note="2 copies 49 mer 91% conserved"
repeat_region
11168..11361
/note="MLT1B repeat: matches 101..309 of consensus"
repeat_region
11448..11736
/note="AluY repeat: matches 1..309 of consensus"
repeat_region
12109..12515
/note="L2 repeat: matches 1747..2155 of consensus"
repeat_region
12798..12947
/note="MIR repeat: matches 29..206 of consensus"
repeat_region
12975..13026
/note="MIR repeat: matches 186..237 of consensus"
repeat_region
13005..13103
/note="L2 repeat: matches 2417..2516 of consensus"
repeat_region
13331..13664
/note="MLT1B repeat: matches 4..361 of consensus"
repeat_region
13704..13907
/note="LIR33 repeat: matches 297..511 of consensus"
repeat_region
15450..15530
/note="L2 repeat: matches 2668..2749 of consensus"
repeat_region
15876..15935
/note="2 copies 30 mer 100% conserved"
repeat_region
16202..16234
/note="MER81 repeat: matches 2..94 of consensus"
misc_feature
17335..17668
/note="match: GSS: Em:AQ104665"
repeat_region
17419..17715
/note="AluSx repeat: matches 1..296 of consensus"
misc_feature
17995..18362
/note="match: GSS: Em:AQ151578"
repeat_region
19374..19496
/note="L2 repeat: matches 2579..2709 of consensus"
repeat_region
20271..20430
/note="MER20 repeat: matches 55..218 of consensus"
repeat_region
21095..21389
/note="AluSx repeat: matches 1..295 of consensus"
repeat_region
21830..21946
/note="MIR repeat: matches 42..152 of consensus"
repeat_region
22328..22772
/note="L2 repeat: matches 2242..2710 of consensus"
repeat_region
22768..22848
/note="L2 repeat: matches 2661..2750 of consensus"
misc_feature
23314..24204
/note="CpG island"
misc_feature
23596..23601
/note="Weak data."
repeat_region
25664..25783
/note="L2 repeat: matches 2588..2741 of consensus"
repeat_region
25786..26131
/note="MER33 repeat: matches 1..324 of consensus"
repeat_region
26858..27183
/note="AluJb repeat: matches 1..301 of consensus"
repeat_region
27208..27334
/note="MIR repeat: matches 81..221 of consensus"
repeat_region
27673..27833
/note="MIR repeat: matches 7..175 of consensus"
misc_feature
complement(28112..28516)
/note="match: GSS: Em:AQ212338"

misc_feature	28535..29088	/note="match: GSS: Em:B51862"	
repeat_region	28560..28694	/note="MIR repeat: matches 79. .210 of consensus"	
misc_feature	28588..28828	/note="match: GSS: Em:B78941"	
misc_feature	28588..28841	/note="match: GSS: Em:B88272"	
repeat_region	28849..29048	/note="MER20 repeat: matches 2. .214 of consensus"	
repeat_region	30148..30232	/note="12 repeat: matches 2610. .2692 of consensus"	
misc_feature	complement(31265..31862)	/note="match: GSS: Em:AQ625926"	
misc_feature	complement(31333..31862)	/note="match: GSS: Em:AQ625934"	
misc_feature	complement(31491..31857)	/note="match: GSS: Em:AQ131221"	
repeat_region	31733..31875	/note="12 repeat: matches 2039. .2197 of consensus"	
repeat_region	33899..33992	/note="12 repeat: matches 2578. .2667 of consensus"	
repeat_region	34111..34437	/note="MER31A repeat: matches 4. .477 of consensus"	
repeat_region	35235..35300	/note="MER5A repeat: matches 121. .189 of consensus"	
repeat_region	35369..35435		
Query Match 7.3%; Score 242; DB 9; Length 132449;			
Best Local Similarity 49.4%; Pred. No. 7.5e-26;			
Matches 719; Conservative 0; Mismatches 725; Indels 12; Gaps 3;			
QY	997 cattaccagtggcagaagaccgcgcgaagtaccagtggaagaattagatgccactccag	1056	
Db	102384 CAGTAGCAGCGCAGTAATAGTAGTGGTAGTAGCAGCAGCTTGACGTAGCTGGCAGGAG	102443	
QY	1057 aggatgacttcgattagattgttacagaatctcccaggagaagtagaattagattagatg	1116	
Db	102444 GAGTAGCAGCAGCAGCAGTGGCCGTAGCTGTGGCAGGAGGATAGCAGCAGCAGCAGT	102503	
QY	1117 aagagcaactgaagaatacaaggaagtgggaccacggaagaagaccacgaag	1176	
Db	102504 CGTAGCTGTGGCAGGAGAGTAGCAGCAGCAGCTTGCCGTAGCTGTGGCAGGAGGAG	102563	
QY	1177 aattagatgcactccagaggatggatttcgattagacgaactgcagaagagaagaaca	1236	
Db	102564 TAGCAGCAGCAGCAGTGGCCG-TAGCTGTGGCAGGAGGAGTAGCAGCAGCAGCAGT	102622	
QY	1237 gaagaacgttagaggagaagaacaagagaagctgcagaagagaagtatcagaagaaa	1296	
Db	102623 GTAGCT----GTGGCAGGAGTAGCAGCAGCAGCTTGCCGTAGCTGTGGCAGGAGGAG	102677	
QY	1297 ctccagaaggaagaagattagaggcaactccagagatgatttcgattagatggaa	1356	
Db	102678 TAGCAGCATCAGCAGTGGCAGTAGCTGTGGCAGGAGAGTAGCAGCATCAGCAGTGGCAG	102737	
QY	1357 ctacattagaagaacgcgaagaactgcagaagagaagaacccgttagaggggagaagaa	1416	
Db	102738 TAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAG	102797	
QY	1417 ccgtagaggagagaacacgcgttagagggagaagaactgcagaagagaagaagattag	1476	
Db	102798 CAGCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCT	102857	
QY	1477 aggcaactccagagattgacttccaattagaagaacccatcaggagaagagaggggaag	1536	
Db	102858 CAGGAGTAGCAGCAGCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAG	102917	
QY	1537 gagaaggaaggggagaagagaagcgttagtagcagtgccagtagtggtggccgaac	1596	
Db	102918 TTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAG	102977	
QY	1597 cggtagaagtagtgcactcgtcagcctgtcaaaccaatggctcgtccaacggcagatg	1656	

Db	102978	GAGTAGCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCAGCAGCTGTTGCAGTAGCTG	103037
QY	1657	aaactttatttcgttgatatcttagataaacgatttaaogtatgcagacattacatcccttg	1716
Db	103038	TGGCAGGAGGAGTAGCAGCAGCAGCAGTTGCCGTAGCTGTGGCAG-----GAGGATTAG	103091
QY	1717	agccattatttaaacaaactcctcaaggatcctgatgcagagagagcgtgttaacagtaccat	1776
Db	103092	CACCATCAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCATCAGTAGTTGCAGTAGCTG	103151
QY	1777	caaaggaagcacctgttacaagtaccagtggcagtagggcccgcgcaagaagtgtccaacgg	1836
Db	103152	TGGCAGCAGGAGTAGCAGCAGCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCAT	103211
QY	1837	aagaattgatgcgaactccaagagagcagatttcgaattagaaggaactgcagaagctccag	1896
Db	103212	CAGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGG	103271
QY	1897	aggaaggagaatttagtattagaaggagaagaaaccaacggaagaagagccaaagaag	1956
Db	103272	CAGGAGTAGCAGCATCAGTAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCAGCAG	1033
QY	1957	gagagccaacagagaagtagtccagaagaagaatttagaggcaactccagaggagcatt	2016
Db	103332	TTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAG	103391
QY	2017	tcgaattagaagaaccaacagaggaagaagtagaagaacccgttagagggagaagaactg	2076
Db	103392	CAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGG	103451
QY	2077	cagaaggagaagaagtgggaagaggtacctgcagaagttagaagaagtgggaaggtacctg	2136
Db	103452	CAGGAGGAGTAGCAGCATCAGCAGTTGCAGTAGCTGTGGCAGGAGGAGTAGCAGCATCAG	103511
QY	2137	cagaagtagaagaagtgggaagaggtaccagaagaagttagaagaggtaccgcgagaagtag	2196
Db	103512	CAGTTGCAGCAGCTGTGGCAGGAGTAGCAGCATCGGCAGTTGCAGCAGCTGTGSCAGGAG	103571
QY	2197	aagaagtgggaaggttaccagaagaagtgggaagggTaccagaagaagtgggaagaggtac	2256
Db	103572	GAGTAGCAGCATCGGCAGTTGCAGCAGCTGTGGCAGGAGGAGTAGCAGCATCGGCAGTTG	103631
QY	2257	cagaagaagtgggaaggttaccagaagaagtgggaagaagtgggaagaagtagaagaagtag	2316
Db	103632	CAGCAGCTGTGGCAGGAGGAGTAGCAGCATCGGCAGTTGCAGCAGCTGTGSCAGGAGGAG	103691
QY	2317	aggtaccagcggtagtagaagtagaagtagaccagcggtagtagaagaagaaggtgccagaag	2376
Db	103692	TAGCAGCATCGGCAGTTGCAGTAGCTGTGGCAGGAGTAGCAGCATCGGCAGTTGCAGTAG	1037
QY	2377	aagttagaagaagaagaagaggaagaaccagtagaggaagaagatgtattacaattag	2436
Db	103752	CTGTGGCAGGAGTAGCAGCAGCGGCAGTTGCAGCAGCTGTGGCAGGAGTAGCAGCAGCAG	103811
QY	2437	taataccatcggaaga	2452
Db	103812	CAGTTGCAGTGCAGGA	103827
RESULT 28			
AC018348/C			
LOCUS			
DEFINITION			
AC018348 Homo sapiens chromosome 15 clone RP11-483E23, WORKING DRAFT			
SEQUENCE, 9 unordered pieces.			
ACCESSION			
AC018348.11 GI:12408462			
VERSION			
KEYWORDS			
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE			
human.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE			
1 (bases 1 to 178022)			

AUTHORS

Abola, A.P., Bruno, D., Conn, L., Dela Rosa, M., Dela Rosa, M., Faulkner, D.,
Federspiel, N., Glukhov, S., Hansen, N., Herman, Z.S., Hyman, R.,
Mao, J., Lam, B., Marathe, R., Miranda, M., Morehouse, A.J., Nguyen, M.,
Oefner, P., Palm, C.J., Ramirez, D., Southwick, A.M., Wilhelmy, J.,
Yu, S. and Davis, R.W.

JOURNAL

REFERENCE

Unpublished
2 (bases 1 to 178022)

AUTHORS

Bruno, D., Conn, L., Dela Rosa, M., Federspiel, N., Foreman, P.,
Glukhov, S., Hansen, N., Hyman, R., Mao, J., Marathe, R.,
Morehouse, A.J., Oefner, P., Palm, C.J., Ramirez, D., Wilhelmy, J.,
Yu, S. and Davis, R.W.

TITLE

JOURNAL

Submitted (09-DEC-1999) DNA Sequencing and Technology Center,
Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA

COMMENT

On Jan 24, 2001 this sequence version replaced gi:12043889.

----- Genome Center

Center: Stanford DNA Sequencing and Technology Development

Center

Center code: SDSTDC

Web site: <http://sequence-www.stanford.edu/group/human/>

Contact: hum-info@sequence.stanford.edu

----- Project Information

Center project name: 732

Center clone name: RP11-483E23

----- Summary Statistics

Sequencing Vector: M13mp18; X02513

Chemistry: Dye-primer; 11% of reads

Assembly program: Phrap; version 0.990319

Consensus quality: 172592 bases at least Q40

Consensus quality: 174564 bases at least Q30

Consensus quality: 175289 bases at least Q20

Insert size: 187151; agarose-fp

Insert size: 177222; sum-of-contigs

Quality coverage: 8.3x in Q20 bases; agarose-fp

Quality coverage: 8.8x in Q20 bases; sum-of-contigs.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 9 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* * * * *

* 1 2877: contig of 2877 bp in length

* 2878 2977: gap of unknown length

* 2978 9281: contig of 6304 bp in length

* 9282 9381: gap of unknown length

* 9382 15085: contig of 5704 bp in length

* 15086 15185: gap of unknown length

* 15186 25650: contig of 10465 bp in length

* 25651 25750: gap of unknown length

* 25751 49398: contig of 23648 bp in length

* 49399 69137: gap of unknown length

* 69138 69237: contig of 19639 bp in length

* 69238 88242: contig of 19005 bp in length

* 88243 88342: gap of unknown length

* 88343 124370: contig of 36028 bp in length

* 124371 124471: gap of unknown length

* 124471 178022: contig of 53552 bp in length.

* Location/Qualifiers

* 1. .178022

* /organism="Homo sapiens"

* /db_xref="taxon:9606"

* /chromosome="15"

* /clone="RP11-483E23"

* /clone_lib="RPCI human BAC library 11"

* 1. .2877

* /note="assembly_name:Contig19"

* 2978. .9281

* /note="assembly_name:Contig20"

misc_feature

misc_feature

misc_feature 9382. .15085
/note="assembly_name:Contig21"
misc_feature 15186. .25650
/note="assembly_name:Contig22"
misc_feature 25751. .49398
/note="assembly_name:Contig23"
misc_feature 49499. .69137
/note="assembly_name:Contig24"
misc_feature 69238. .88242
/note="assembly_name:Contig25"
misc_feature 88343. .124370
/note="assembly_name:Contig26"
misc_feature 124471. .178022
/note="assembly_name:Contig27"

BASE COUNT 51863 a 38471 c 37841 g 49044 t 803 others

ORIGIN

Query Match 7.3%; Score 242; DB 2; Length 178022;

Best Local Similarity 50.1%; Pred. No. 7.4e-26;

Matches 691; Conservative 0; Mismatches 675; Indels 14; Gaps 3;

QY 1056 gaggatgacttcgattgattacagaatctcccgaggaagtagattagattagat 1115
DB 149970 GCGGAGCAGGAGGAGATGATGCGGAGAGAGGAGGAGAGATGCGGAGGAGGAGAT 149911
QY 1116 gaagaggaactgaagaagaatcaacggaagtgggaccaaaggaggaagaccaccaa 1175
DB 149910 GATGTGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 149851
QY 1176 gaattagatgccactccagagagattgattcgattagacgaaactcgagaagagaaac 1235
DB 149850 GATACGGGAGCAGGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149792
QY 1236 agaagaac-----gtagagggagaagaacacagaagaagctgcagaagagaatcaga 1291
DB 149791 GGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149732
QY 1292 agaactccagaagagaagaagattagaggcaactccagagagatttcgcattaga 1351
DB 149731 AGATACGGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149672
QY 1352 tggaactacattagaagaaccgaagaactgcagaagaggaagaaccgttagagggaga 1411
DB 149671 AGATACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149612
QY 1412 agaaccgttagagggagaagaacccttagaggagaagaagctgcagaagagaagaaga 1471
DB 149611 TAAGGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149552
QY 1472 gttagaggaactccagagagattcctccattagaagaaccatcagagaagagaagg 1531
DB 149551 TACGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149492
QY 1532 ggaagagaaggaaggggaaggaaggaaggaagcgttagtagcagtcagctagtgcc 1591
DB 149491 AGATACGGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149432
QY 1592 cgaaccggttagaagtagtagtgcactcctcagcctgcacaaaccaatggctcccaacggc 1651
DB 149431 AGAAGATGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149372
QY 1652 agatgaactttattcgttgatatcttagataaacgatttaacgtatgcacattacatc 1711
DB 149371 AGAAGATGTAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149312
QY 1712 ctttagccattattaaacaaatcctcaaggatccttgatgcaggagaggtcttaacagt 1771
DB 149311 AGATGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 149252
QY 1772 accatcaaggaagaccctgtacaagtaccagttggcagtaggggcccgcgcaagaagtgc 1831


```
QY 1352 tggaaactacattagaagaacccagaactgcagaagggagaagaacccgtagaggaga 1411
    || || || || || || || || || || || || || || || || || || || || ||
Db 70716 AGATACGGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70657
QY 1412 agaaccgttagaggagaagaacccgttagaggagaagaacccgttagaggagaaga 1471
    || || || || || || || || || || || || || || || || || || || || ||
Db 70656 TAAGGGAGCAGGAGGAGATATGGAGGCAAGAGGAGAGATATGCACGAGCAGGAGGA 70597
QY 1472 gttagaggcaactcagagagatgacttccattagaagaaccatcaggagaagaagg 1531
    || || || || || || || || || || || || || || || || || || || || ||
Db 70596 TACGGAAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70537
QY 1532 ggaaggagaaggagaaggagaagagagagagagagagagagagagagagagagag 1591
    || || || || || || || || || || || || || || || || || || || || ||
Db 70536 AGATACGGGAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70477
QY 1592 cgaaccggttagagtagtgcactcgtcagcctgtcgaacccaatggtcgtccaaagg 1651
    || || || || || || || || || || || || || || || || || || || || ||
Db 70476 AGAAGATGTGGAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70417
QY 1652 agatgaacttttctgtatctcttagataacgatttaacgtatgcacattacatc 1711
    || || || || || || || || || || || || || || || || || || || || ||
Db 70416 AGAAGATGTGGAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70357
QY 1712 ctttgagccattattaaacaaatcctcaaggatcctgtgcaggagaggtcttaacagt 1771
    || || || || || || || || || || || || || || || || || || || || ||
Db 70356 AGATGTGGAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70297
QY 1772 accataaagaagaccctgtacaagtaccagtggcagtagggccgcgcaagaagtgc 1831
    || || || || || || || || || || || || || || || || || || || || ||
Db 70296 AGAAGATACGGGAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70237
QY 1832 aacggaagaattgatgcaactccaagaaggacgatttctgaattagaagaaactcgaagc 1891
    || || || || || || || || || || || || || || || || || || || || ||
Db 70236 AGGAGAAGATAAGGAGCAGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 70177
QY 1892 tccagaaggagaaggaatttagtagaagagagagagagagagagagagagagagag 1951
    || || || || || || || || || || || || || || || || || || || || ||
Db 70176 AGGAGAAGATACGAGCAGGAGGAGAGATGTCAGGAACAGGAGAGAGAGAGAGAG 70120
QY 1952 agaagagagcccaagaaggaagtgccagaagaagaattagagcaactccaagga 2011
    || || || || || || || || || || || || || || || || || || || || ||
Db 70121 ----GGAGCAGGAGAGAGAGATCAGAACAGGAGAGAGATGCGGAGCAGGAGGAGA 70066
QY 2012 cgatttcgaattagaagaaccaagagagagagagagagagagagagagagagagag 2071
    || || || || || || || || || || || || || || || || || || || || ||
Db 70065 TAAGGGAGCAGGAGGAGAGAGATACGGGAGCAGAGAGAGAGAGATACGGGAGCAGGAGGAGA 70006
QY 2072 aactgcagaaggagaagaagtgaagaaggtacctgcagaagtagagaagaagtggaaaggt 2131
    || || || || || || || || || || || || || || || || || || || || ||
Db 70005 AGATATGGGAGCAGGAGGAGAGAGATACGAGAGCAGGAGGAGAGATGTCAGGAACAGGAGAG 69946
QY 2132 acctgcagaagtagaagaagtgaagaaggtaccagaagaagtagaagaagtagtaccgcaga 2191
    || || || || || || || || || || || || || || || || || || || || ||
Db 69945 AGAAGATGGGGAGCAGGAGGAGAGAGATGTGGAGCAGGAGGAGAGAGAGATGCAGAGAACAGG 69886
QY 2192 agtagaagaagtgaagaaggtaccagaagaagtggaagaggtaccagaagaagtggaaga 2251
    || || || || || || || || || || || || || || || || || || || || ||
Db 69885 AGGAGAAGATGCGGAGCAGGAGGAGAGAGATAGGGAGCAGGAGGAGAGAGAGATACGGGAGC 69826
QY 2252 ggtaccagaagaagtggaagaaggtaccagaagaagtggaagaagtggaagaagtagaaga 2311
    || || || || || || || || || || || || || || || || || || || || ||
Db 69825 AGGAGGAGAGAGATACGAGAGCAGGAGAGAGATGTGCAGGAACAGGAGAGAGAGATGGGGG 69766
QY 2312 agtagaggtaccagcggttagaagtagaagtagaccagcggttagaagaagaaggtgcc 2371
    || || || || || || || || || || || || || || || || || || || || ||
Db 69765 AGCAGGAGGAGAGATGTGTGACAGAGAGAGAGAGATGCAGAGACAGGAGGAGAGAGATGC 69706
QY 2372 agaagaagtagaagaagaagaagagagagagagagagagagagagagagagagagagag 2431
    || || || || || || || || || || || || || || || || || || || || ||
Db 69705 GGAGGCAGGAGGAGAGATAGGGAGCAGGAGGAGAGAGATACGGGAGCAGGAGGAGAGAGA 69646
```

```
RESULT 30
AC068037
LOCUS 206136 bp DNA linear PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-757K22 from 4, complete sequence.
AC068037
VERSION AC068037.5 GI:18042511
KEYWORDS HTG.
SOURCE human.
```

```
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE
1 (bases 1 to 206136)
```

```
AUTHORS
Tulston, J.E. and Waterston, R.
```

```
JOURNAL
Genome Res. 8 (11), 1097-1108 (1998)
```

```
MEDLINE
99063792
```

```
REFERENCE
2 (bases 1 to 206136)
```

```
AUTHORS
Isak, A., Kozlowski, A. and Creason, K.
```

```
JOURNAL
The sequence of Homo sapiens BAC clone RP11-757K22
```

```
REFERENCE
3 (bases 1 to 206136)
```

```
AUTHORS
Waterston, R.H.
```

```
JOURNAL
Direct Submission
```

```
REFERENCE
4 (bases 1 to 206136)
```

```
AUTHORS
Submitted (27-APR-2000) Genome Sequencing Center, Washington
```

```
JOURNAL
University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
```

```
REFERENCE
5 (bases 1 to 206136)
```

```
AUTHORS
Waterston, R.H.
```

```
JOURNAL
Direct Submission
```

```
REFERENCE
Submitted (03-JAN-2002) Genome Sequencing Center, Washington
```

```
JOURNAL
University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
```

```
COMMENT
On Jan 3, 2002 this sequence version replaced gi:15799672.
```

```
Center: Washington University Genome Sequencing Center
```

```
Center code: WUGSC
```

```
Web site: http://genome.wustl.edu/gsc
```

```
Contact: sapiens@watson.wustl.edu
```

```
----- Summary Statistics
```

```
Center project name: H_NH0757K22
```

```
-----
```

```
NOTICE: This sequence may not represent the entire insert of this
```

```
clone. It may be shorter because we only sequence overlapping
```

```
clone sections once, or longer because we provide a small overlap
```

```
between neighboring data submissions.
```

```
This sequence was finished as follows unless otherwise noted:
```

```
all regions were double stranded, sequenced with an alternate
```

```
chemistry, or covered by high quality data (i.e., phred quality >=
```

```
30); an attempt was made to resolve all sequencing problems, such
```

```
as compressions and repeats; all regions were covered by sequence
```

```
from more than one subclone; and the assembly was confirmed by
```

```
restriction digest.
```

```
MAPPING INFORMATION:
```

```
Mapping information for this clone was provided by Dr. John D.
```

```
McPherson, Department of Genetics, Washington University, St. Louis
```

```
MO. For additional information about the map position of this
```

```
sequence, see http://genome.wustl.edu/gsc
```

```
SOURCE INFORMATION:
```

```
The RPCI-11 human BAC library was made from the blood of one male
```

```
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
```

```
Tateno, M., Catanese, J. and de Jong, P.J. (1998) An improved
```

```
approach for construction of bacterial artificial chromosome
```

```
libraries. Genomics 51:1-8. The clone may be obtained either from
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 15:29:46 ; Search time 2486.21 Seconds
(without alignments)
18115.662 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagtaacaa.....tctataataataataattc 3337

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 13736207 seqs, 6748477542 residues

Word size : 0
Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	423	12.7	633	12	AZ569608 263PVD09
2	34	1.0	337	12	AZ852656
3	33	1.0	825	12	AZ682647
4	32	1.0	189	10	BF854267
5	32	1.0	229	10	BF854267
6	32	1.0	250	10	BI708633
7	32	1.0	380	9	AW101380
8	32	1.0	905	12	AZ677174
9	31	0.9	774	10	BG618297
10	30	0.9	144	10	BG959271
11	30	0.9	166	9	BE146466
12	30	0.9	208	10	T05426
13	30	0.9	255	9	AW886018
14	30	0.9	304	9	AW619681
15	30	0.9	335	9	AU097484
16	30	0.9	362	9	AI650676
17	30	0.9	367	10	BM169373

18	30	0.9	383	12	AZ569608
19	30	0.9	396	10	BG683139
20	30	0.9	403	10	BM034050
21	30	0.9	586	10	BI284201
22	30	0.9	627	12	AZ836143
23	30	0.9	785	12	BI150657
24	30	0.9	824	10	BE488177
25	30	0.9	835	12	BI431115
26	30	0.9	1087	12	AQ752154
27	30	0.9	1120	10	BM464445
28	29	0.9	30	12	TA247F06P
29	29	0.9	34	12	TA222B09P
30	29	0.9	68	12	CNS048HA
31	29	0.9	68	12	CNS048HA
32	29	0.9	104	9	AW885174
33	29	0.9	112	9	AL641523
34	29	0.9	113	12	BH500312
35	29	0.9	119	9	AW860980
36	29	0.9	122	10	BE974031
37	29	0.9	133	10	BG995888
38	29	0.9	134	9	AW071377
39	29	0.9	135	9	AW889107
40	29	0.9	138	9	AA790025
41	29	0.9	143	9	AW345955
42	29	0.9	148	10	BF831412
43	29	0.9	149	10	BG981647
44	29	0.9	150	10	BF924884
45	29	0.9	151	10	BG736025

ALIGNMENTS

RESULT 1
AZ569608
LOCUS 633 bp DNA linear GSS 15-MAY-2001
DEFINITION 263PVD09 Pv MBN #30 Plasmodium vivax genomic 3', DNA sequence.
ACCESSION AZ569608
VERSION AZ569608.1 GI:13979872
KEYWORDS GSS.
SOURCE malaria parasite P. vivax.
ORGANISM Plasmodium vivax
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 633)
AUTHORS Carlton, J.N.-R. and Dame, J.B.
TITLE The Plasmodium vivax and P. berghei gene sequence tag projects
JOURNAL Parasitol. Today 16 (10), 409 (2000)
COMMENT Contact: Dame JB
Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

FEATURES
Location/Qualifiers
1..633
/organism="Plasmodium vivax"
/strain="Salvador I (Collins, W. 1972. J. Parasitol. 69, 497-598)"
/db_xref="taxon:5855"
/clone_lib="pv MBN #30"
/dev_stage="asexual blood forms"
/lab_host="Salmirali boliviensis"
/note="Vector: pBluescript SK(+) vector DNA, phagemid excised from lambda ZAP; Site_1: EcoR V; Site_2: EcoR V; Host leukocytes were extracted from P. vivax infected blood using the following methods: first, infected blood was activated by the addition of 0.5 ml of ADP (40mg/ml) per 10 ml blood. Then blood was passed over a column of acid washed 0.1 mm glass beads, then through a Plasmidpur filter, followed by passage through a column of pre-wet

Email: asimpson@ludwig.org.br.
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL

```
Site_1: DraIII(Y); Site_
```



```

RESULT 10
BG959271/c
LOCUS
DEFINITION CM2-CT0781-190301-701-a01_1 CT0781 Homo sapiens cDNA, mRNA
sequence.
ACCESSION BG959271
VERSION BG959271.1 GI:14377442
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 144)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2at2-CM2-CT0781-
190301-701-a01_1&t3=2001-03-19&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 118.
Location/Qualifiers
1..144
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="CT0781"
/dev_stage="Adult"
/note="Organ: colon; Vector: puc18; Site_1: SmaI; Site_2:
SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
,716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."
BASE COUNT 78 a 29 c 23 g 14 t
ORIGIN
Query Match 0.9%; Score 30; DB 10; Length 144;
Best Local Similarity 100.0%; Pred. No. 1.3e+03;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 192 gaaataattttttttttttttttttttttt 221
|||||
Db 90 GAAAAAATTTTTTTTTTTTTTTTTTTTTT 61

RESULT 11
BE146466/c
LOCUS
DEFINITION QV0-HT0216-011199-043-b06 HT0216 Homo sapiens cDNA, mRNA sequence.
ACCESSION BE146466
VERSION BE146466.1 GI:8609190
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 166)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2-QV0-HT0216-011
199-043-b06&t3=1999-11-01&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 85.
Location/Qualifiers
1..166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="HT0216"
/dev_stage="Adult"
/note="Organ: head_neck; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 61 a 27 c 42 g 36 t
ORIGIN
Query Match 0.9%; Score 30; DB 9; Length 166;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps
QY 194 aaaaaattttttttttttttttttttttt 223
|||||
Db 84 AAAAAATTTTTTTTTTTTTTTTTTTTTTCT 55

RESULT 12
T05426/c
LOCUS
DEFINITION T05426 Fetal brain, Stratagene (cat#936206) Homo sapiens cDNA
clone HFCW86 similar to EST containing Alu repeat, mRNA sequence.
ACCESSION T05426
VERSION T05426.1 GI:316578
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 208)
Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
3,400 expressed sequence tags identify diversity of transcripts
from human brain
Nature Genet. 4, 256-267 (1993)
JOURNAL 93364420
MEDLINE
CONTACT: Adams, MD

```

The Institute for Genomic Research
932 Clopper Road, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423
Email: mdadams@tigr.org
Seq primer: M13-21.

FEATURES

source
1. .208
/organism="Homo sapiens"
/db_xref="ATCC (inhost):82144"
/db_xref="taxon:9606"
/clone_lib="HFCW86"
/note="Vector: Fetal brain, Strategene (cat#936206),
LambdazAP-II; 17-18 wk gestation, female;
oligo-dT + random primed cDNA synthesis; LambdazAP-II
vector, 1.0kb average inser size." 3 others

BASE COUNT

71 a 37 c 53 g 44 t
ORIGIN

Query Match

Best Local Similarity 0.9%; Score 30; DB 10; Length 208;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

193 aaaaaattttttttttttttttttttttttg 222
|||||

Db

200 AAAAAAATTTTTTTTTTTTTTTTTTTTGT 171

RESULT 13

AW886018/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 255)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=RCL-OT0075-120
400-011-h07&t3=2000-04-12&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 41
High quality sequence stop: 254.

FEATURES

source
1. .255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="OT0075"
/dev_stage="Adult"
/note="Organ: ovary; Vector: puc18; Site1: SmaI; Site_2:

SmaI; A mini-library was made by cloning products derived
from ORESTES PCR (U.S. Letters Patent application No. 196
716 - Ludwig Institute for Cancer Research) profiles
into the pUC 18 vector. Reverse transcription of tissue
mRNA and cDNA amplification were performed under low
stringency conditions."

BASE COUNT

102 a 50 c 37 g 66 t
ORIGIN

Query Match

Best Local Similarity 0.9%; Score 30; DB 9; Length 255;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

192 gaaaaattttttttttttttttttttttttt 221
|||||

Db

192 GAAAAAATTTTTTTTTTTTTTTTTTTT 163

RESULT 14

AW619681/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT

Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
1 (bases 1 to 304)
Smith,T.P.L., Fahrenkrug,S.C., Rohrer,G.A., Simmen,F.A., Rexroad
C.E. and Keefe,J.W.
Mapping of expressed sequence tags from a porcine early embryonic
cDNA library
Anim. Genet. 32 (2), 66-72 (2001)
21314990
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: GGAAACAGCTATGACCATG
BACKWARD: GTAAACGACGCCAGT
Seq primer: AATTACCCCTCACTAAAGG.

FEATURES

source
1. .304
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC PBE"
/tissue_type="Day 12 whole embryos"
/lab_host="XLOLR"
/note="Vector: pBLUESCRIPT SK-; Site_1: EcoRI; Site_2:
XhoI; Library made from pool of embryos in spherical and
filamentous stages of development (7.5% and 92.5%,
respectively, of each stage) as described in Choi et al,
Endocrinology 137, 1457-67, 1996."

BASE COUNT

107 a 64 c 81 g 52 t
ORIGIN

Query Match

Best Local Similarity 0.9%; Score 30; DB 9; Length 304;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

194 aaaaaatttttttttttttttttttttttgt 223
|||||

Db

254 AAAAAAATTTTTTTTTTTTTTTTTTTTGT 225

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cbyJ mice infected with pY17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT 189 a 32 c 42 g 104 t
ORIGIN

Query Match 0.9%; Score 30; DB 10; Length 367;
Best Local Similarity 100.0%; Pred.No.5e+02; 0; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;

QY 192 gaaaaaatTTTTTTTTTTTTTTTTTTTTT 221
|||||TTTTTTTTTTTTTTTTTTTTTTTTTTTT
Db 86 GAAAAAATTTTTTTTTTTTTTTTTTTTTT 57

RESULT 18

AZ656316 383 bp DNA linear GSS 14-DEC-2000
LOCUS 1M0531117R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0531117 R, DNA sequence.

ACCESSION AZ656316
VERSION
KEYWORDS GSS.
SOURCE house mouse.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly,
M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D.,Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts

TITLE

Unpublished (2000)

JOURNAL

COMMENT Contact: Robert B. Weiss

University of Utah Genome Center

Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT

84112, USA

Tel: 801 585 5606

Fax: 801 585 7177

Email: ddunn@genetics.utah.edu

Insert Length: 10000 Std Error: 0.00

Plate: 0531 row: I column: 17

Seq primer: CACACAGGAACACGCTATGACC

Class: plasmid ends

High quality sequence stop: 383.

Location/Qualifiers

FEATURES

source

1. .383
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0531117"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.

musculus C57BL/6J (male) was obtained from the Jackson Laboratory Mouse DNA Resource (<http://www.jax.org/resources/documents/dnares/>). The DNA was hydrodynamically sheared by repeated passage through a 0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adaptor DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1), a copy-number inducible derivative of plasmid R1. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adaptor mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent E. coli XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 167 a 36 c 117 g 63 t
ORIGIN

Query Match 0.9%; Score 30; DB 12; Length 383;
Best Local Similarity 100.0%; Pred.No.4.8e+02; 0; Indels 0; Gaps 0;
Matches 30; Conservative 0; Mismatches 0;

QY 1520 aagaagagaagggaaggaaggaaggaagg 1549
|||||TTTTTTTTTTTTTTTTTTTTTTTTTTTT
Db 68 AGAAGGAGAGGGGAGGAGGAGGAGGAGG 97

RESULT 19

BG683139 396 bp mRNA linear EST 01-MAY-2001
LOCUS 602851625F1 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:4761511 5',
DEFINITION mRNA sequence.

ACCESSION BG683139
VERSION
KEYWORDS EST.
SOURCE human.

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 396)

NIH-MGC <http://imgc.ncbi.nih.gov/>.

National Institutes of Health, Mammalian Gene Collection (MGC)

TITLE

Unpublished (1999)

JOURNAL

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgabbs-r@mail.nih.gov

Tissue procurement: ATCC

cDNA Library Preparation: Ling Hong/Rubin Laboratory

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.

Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

<http://image.llnl.gov>

Plate: LLCM1614 row: b column: 08

High quality sequence stop: 115.

Location/Qualifiers

FEATURES

source

1. .396
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4761511"
/clone_lib="NIH_MGC_47"
/tissue_type="neuroblastoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: brain; Vector: pOTB7; Site:1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit

(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC Library."

BASE COUNT
ORIGIN

Query Match 0.9%; Score 30; DB 10; Length 396;
Best Local Similarity 100.0%; Pred. No. 4.7e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 193 aaaaaattttttttttttttttttttt 222
|||||
Db 88 AAAAAAATTTTTTTTTTTTTTTTTTTG 117

RESULT 20
BM034050/c

LOCUS
DEFINITION
kh72c01.y1 Ascaris suum female gonad DZ pamp1 v2 Chiapelli McCarter
PRECURSOR ; , mRNA sequence.

ACCESSION
VERSION
BM034050
EST.

KEYWORDS
SOURCE
pig roundworm.

ORGANISM

Ascaris suum
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.

REFERENCE

AUTHORS
McCarte J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
Gibbons, M., Kitter, E., Bennett, J., Franklin, C., Tsagarelshvili, R.,
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R. and
Wilson, R.

TITLE
The Washington Univ. Nematode EST Project, 1999

JOURNAL

COMMENT
Unpublished (1999)

Contact: McCarter JP

The Washington Univ. Nematode EST Project, 1999

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

High quality sequence stop: 115.

FEATURES

source

1. .403

Location/Qualifiers

/organism="Ascaris suum"

/db_xref="taxon:6253"

/clone_lib="Ascaris suum female gonad DZ pamp1 v2

Chiapelli McCarter"

/sex="Female"

/tissue_type="Dissected female gonad (DZ-differentiation

zone, middle region of gonad)"

/dev_stage="Adult"

/lab_host="DH10b"

/note="Vector: pamp1 (Gibco); Site.1: NotI; Site.2: SalI;

The library was constructed by Brandi Chiapelli and Dr.

James McCarter at Washington University, St. Louis. The

cDNA was made by using Dynabead oligo-dT priming (Dynal).

PCR based library using a modified protocol from the

SMART PCR cDNA Synthesis Kit from Clontech. Directionally

cloned into the UDG sites of pamp1. Dissected nematode

tissues were provided by Dr. Alan Scott (ascott@jhsph.edu

) of the School of Public Hygiene and Public Health at

John Hopkins University in Baltimore, MD."

227 a 43 c 82 g 51 t

BASE COUNT
ORIGIN

Query Match 0.9%; Score 30; DB 10; Length 403;
Best Local Similarity 100.0%; Pred. No. 4.6e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 gaaaaattttttttttttttttttttt 221
|||||
Db 178 GAAAAAATTTTTTTTTTTTTTTTTTTT 149

RESULT 21
BI284201/c

LOCUS

DEFINITION

UI-R-CX0s-ccn-d-08-0-UI-s1 UI-R-CX0s Rattus norvegicus cDNA clone

EST.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus

1 (bases 1 to 586)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

Contact: Soares, MB

Program for Rat Gene Discovery and Mapping

University of Iowa

451 Eckstein Medical Research Building Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: msoares@blue.weeg.uiowa.edu

The sequence contained an oligo-dT track that was present in the

oligonucleotide that was used to prime the synthesis of first

strand cDNA and therefore this may represent a bonafide poly A

tail. The sequence tag present in the cDNA between the NotI site

and the oligo-dT track served to identify it as a clone from the

non-normalized rat placenta pool library cDNA library preparation:

M.B. Soares Lab Clone distribution: clones will be available

through Research Genetics (www.resgen.com) the following repetitive

elements were found in this cDNA sequence: 40-176,

>POLY_A#Simple_repeat 206-228, >GC_rich#Low_complexity 333-353,

>AT_rich#Low_complexity

Seq primer: M13 Forward

POLYA=yes

FEATURES

source

1. .586

Location/Qualifiers

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone_lib="UI-R-CX0s-ccn-d-08-0-UI"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site.1: Not I; Site.2: Eco RI; The UI-R-CX0s

library is a non-normalized library constructed from the

following rat placenta tissues: embryonic day 17,

embryonic day 19, embryonic day 21. For a detailed

description of the library from which this clone was

derived, please visit our web site at

ratseq.eng.uiowa.edu. The subtraction has been previously

described in (Bonaldo, Lennon and Soares, Genome Research

6:791-806, 1996)

TAG_LIB=UI-R-CX0s

TAG_TISSUE=rat placenta pool

TAG_SEQ=TCACGACAGT"

263 a 104 c 143 g 76 t

BASE COUNT
ORIGIN


```
Query Match      0.9%; Score 30; DB 10; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttg 222
|||||
Db 68 AAAAAATTTTttttttttttttttttttttttttttttg 39

RESULT 22
AZ836143
LOCUS      627 bp      DNA      linear      GSS 20-FEB-2001
DEFINITION 2M0130N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0130N15 R, DNA sequence.
ACCESSION  AZ836143
VERSION     AZ836143.1  GI:13006051
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 627)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0130 row: N column: 15
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 627.
FEATURES
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0130N15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="F. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
258 a      82 c      176 g      111 t

Query Match      0.9%; Score 30; DB 10; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttg 222
|||||
Db 68 AAAAAATTTTttttttttttttttttttttttttttttg 39

RESULT 22
AZ836143
LOCUS      627 bp      DNA      linear      GSS 20-FEB-2001
DEFINITION 2M0130N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0130N15 R, DNA sequence.
ACCESSION  AZ836143
VERSION     AZ836143.1  GI:13006051
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 627)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0130 row: N column: 15
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 627.
FEATURES
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0130N15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="F. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
258 a      82 c      176 g      111 t

Query Match      0.9%; Score 30; DB 10; Length 586;
Best Local Similarity 100.0%; Pred. No. 3.2e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttg 222
|||||
Db 68 AAAAAATTTTttttttttttttttttttttttttttttg 39

RESULT 22
AZ836143
LOCUS      627 bp      DNA      linear      GSS 20-FEB-2001
DEFINITION 2M0130N15R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC2M0130N15 R, DNA sequence.
ACCESSION  AZ836143
VERSION     AZ836143.1  GI:13006051
KEYWORDS    GSS.
SOURCE      house mouse.
ORGANISM    Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 627)
AUTHORS     Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
TITLE       Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
JOURNAL     Unpublished (2000)
COMMENT     Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0130 row: N column: 15
Seq primer: CACACAGGAAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 627.
FEATURES
Location/Qualifiers
1..627
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC2M0130N15"
/clone_lib="Mouse 10kb plasmid UUGC1M library"
/sex="Male"
/lab_host="F. Coli strain XL10-Gold, Tl-resistant, F-"
/note="Vector: PWD42nv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adaptored DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative
of pWD42 (gi14732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adaptored mouse DNA was annealed to
adaptored vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."
258 a      82 c      176 g      111 t
```

ORIGIN

```
Query Match      0.9%; Score 30; DB 12; Length 627;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1520 agaagagaaggggaaggagaaggagaagg 1549
|||||
Db 15 AGAAGGAGAGGGGAGGAGAGGAGGAGG 44

RESULT 23
BH150657
LOCUS      785 bp      DNA      linear      GSS 27-AUG-2001
DEFINITION ENPK51TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION  BH150657
VERSION     BH150657.1  GI:15312424
KEYWORDS    GSS.
SOURCE      Entamoeba histolytica.
ORGANISM    Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1 (bases 1 to 785)
AUTHORS     Loftus,B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE       Determination of clone end sequences from Entamoeba histolytica
HMI-IMSS sheared DNA library (2001)
JOURNAL     Unpublished (2001)
COMMENT     Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igr.org
Clones are derived from the Entamoeba histolytica HMI-IMSS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 17
High quality sequence stop: 325.
FEATURES
Location/Qualifiers
1..785
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site_1: Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
205 a      156 c      170 g      254 t

BASE COUNT      205 a      156 c      170 g      254 t
ORIGIN

Query Match      0.9%; Score 30; DB 12; Length 785;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttg 222
|||||
Db 724 AAAAAATTTTttttttttttttttttttttttttttttg 753
```

ORIGIN


```

BASE COUNT      415 a      240 c      295 g      84 t      53 others
ORIGIN

Query Match      0.9%; Score 30; DB 12; Length 1087;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 gaaaaaattttttttttttttttttttttt 221
|||||
Db 1049 GAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 1020

RESULT 27
BM464445/c
LOCUS      BM464445
DEFINITION AGENCOURT_6438703 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:5535531
5', mRNA sequence.
ACCESSION  BM464445
VERSION     BM464445.1 GI:18513487
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1120)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-remail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM1223 row: m column: 04
          High quality sequence start: 3
          High quality sequence stop: 260.
          Location/Qualifiers
            1..1120
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="IMAGE:5535531"
              /clone_lib="NIH_MGC_71"
              /tissue_type="leiomysarcoma"
              /lab_host="DH10B (phage-resistant)"
              /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
              Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
              Average insert size 2.1 kb."
              261 a      342 c      315 g      200 t
              2 others
BASE COUNT      261 a      342 c      315 g      200 t
ORIGIN

Query Match      0.9%; Score 30; DB 10; Length 1120;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 gaaaaaattttttttttttttttttttttt 221
|||||
Db 798 GAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 769

RESULT 28
TA247F06P
LOCUS      TA247F06P
DEFINITION T. brucei sheared genomic DNA clone 247f06, forward sequence,
genomic survey sequence.
ACCESSION  AL483252
VERSION     AL483252.1 GI:11848928
KEYWORDS   GSS.

Trypanosoma brucei.
Trypanosoma brucei
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 30)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barrell, Oxford University Press, 1999).
Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.
Location/Qualifiers
  1..30
    /organism="Trypanosoma brucei"
    /strain="TREU927"
    /db_xref="taxon:5691"
    /clone="247f06"
  8 a      0 c      0 g      22 t
BASE COUNT      8 a      0 c      0 g      22 t
ORIGIN

Query Match      0.9%; Score 29; DB 12; Length 30;
Best Local Similarity 100.0%; Pred. No. 1e+04;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 aaaaaattttttttttttttttttttttt 221
|||||
Db 2 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 30

RESULT 29
TA222B09P
LOCUS      TA222B09P
DEFINITION T. brucei sheared genomic DNA clone 222b09, forward sequence,
genomic survey sequence.
ACCESSION  AL480722
VERSION     AL480722.1 GI:11846491
KEYWORDS   GSS.
Trypanosoma brucei.
Trypanosoma brucei.
Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
Trypanosoma.
1 (bases 1 to 34)
Hall, N., Bowman, S., Lennard, N.J., Doggett, J., Atkin, R.,
Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L.,
Melville, S.E., Rajandream, M.A. and Barrell, B.G.
Direct Submission
Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing
project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton,
Cambridge CB10 1SA, E-mail: barrell@sanger.ac.uk and
nh@sanger.ac.uk
Constructed at the Institute for Genomic Research (TIGR),
Rockville, MD. Genomic DNA isolated from a cloned population of
Trypanosoma brucei (TREU927/4 GUTat 10.1) was mechanically sheared
to give a tight size distribution (
4 kb). The v + i method used for the library construction is
described in detail in Smith, H. and Venter, J.C. (Making small
insert libraries for whole genome shotgun sequencing projects. In
Genome Sequencing: A Practical Approach, eds. M. Vaudin and B.

```

Barrell, Oxford University Press, 1999).

Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available
at http://www.sanger.ac.uk/Projects/T_brucei/.

FEATURES

source
1..34
/organism="Trypanosoma brucei"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="222b09"

BASE COUNT 12 a 0 c 0 g 22 t
ORIGIN

Query Match 0.9%; Score 29; DB 12; Length 34;

Best Local Similarity 100.0%; Pred. No. 9e+03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttt 221
|||||
Db 6 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 34

RESULT 30

CNS048HA

LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
090014 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION

AL279271.1 GI:8016612

VERSION
GSS: genome survey sequence.

KEYWORDS
Tetraodon nigroviridis.

SOURCE

ORGANISM
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.

REFERENCE

AUTHORS
1 (bases 1 to 68)
Roest-Crollius, H., Jaillon, O., Dasilva, C., Fizames, C., Fisher, C.,
Bouneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and
Weissenbach, J.

TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL

REFERENCE
2 (bases 1 to 68)

AUTHORS
Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fizames, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.

TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL

REFERENCE
3 (bases 1 to 68)

AUTHORS
Genoscope.

TITLE
Direct Submission

JOURNAL
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

COMMENT
This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
<http://www.genoscope.cns.fr/Tetraodon>.

FEATURES

source
1..68
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="090014"
/clone_lib="G"
/note="Genoscope sequence ID : COBG090BH07SP1-end ;
PUC-Ori"

BASE COUNT 29 a 2 c 1 g 35 t 1 others
ORIGIN

Query Match

0.9%; Score 29; DB 12; Length 68;

Best Local Similarity 100.0%; Pred. No. 4.5e+03;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaattttttttttttttttttttttttttttttt 221
|||||
Db 28 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 56

Search completed: June 14, 2002, 19:14:07
Job time: 13461 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 17:26:02 ; Search time 352.06 Seconds
(without alignments)
16273.770 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gaattccgtaagaagtaacaa.....tctataaaataataatc 3337

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 1736436 seqs, 858457221 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	17 AAT34620	P. vivax ESP-1 blo
2	3337	100.0	3337	20 AAX15174	DNA encoding a sec
3	3337	100.0	3337	22 AAH76457	Plasmodium vivax E
4	33	1.0	6862	22 AAS46299	Tumour suppressor
5	33	1.0	6862	24 ABL32222	Human immune syste
6	33	1.0	6862	24 AAS61081	Human gene regulat
7	30	0.9	334	22 AAI82170	Human polynucleoti
8	30	0.9	337	22 AAL11387	Human breast cance
9	30	0.9	363	22 AAL20286	Human breast cance

c 10	30	0.9	392	22	AAI90039	Human polynucleoti
c 11	30	0.9	5856	23	ABL09674	Drosophila melanog
c 12	30	0.9	8801	22	AAS45437	Chemically pretrea
c 13	30	0.9	8801	24	ABL33741	Human immune syste
c 14	29	0.9	32	14	AAQ43973	Triple helix formi
c 15	29	0.9	40	13	AAQ25032	Oligonucleotide sp
c 16	29	0.9	325	22	AAI89791	Human polynucleoti
c 17	29	0.9	338	22	AAI80023	Human polynucleoti
c 18	29	0.9	344	22	AAI85033	Human polynucleoti
c 19	29	0.9	352	22	AAI91956	Human polynucleoti
c 20	29	0.9	361	22	AAI84186	Human polynucleoti
c 21	29	0.9	369	22	AAI84071	Human polynucleoti
c 22	29	0.9	404	22	AAI82188	Human polynucleoti
c 23	29	0.9	406	22	AAI87091	Human polynucleoti
c 24	29	0.9	416	22	AAI87993	Human polynucleoti
c 25	29	0.9	417	22	AAI84767	Human polynucleoti
c 26	29	0.9	422	22	AAI89825	Human polynucleoti
c 27	29	0.9	425	22	AAI84618	Human polynucleoti
c 28	29	0.9	467	22	AAK63939	Human immune/haema
c 29	29	0.9	608	22	AAH00504	Trichophyton menta
c 30	29	0.9	758	22	AAI88874	Human polynucleoti
c 31	29	0.9	767	22	AAI88875	Human polynucleoti
c 32	29	0.9	821	21	AAA02473	Human colon.cancer
c 33	29	0.9	910	24	ABL34307	Human immune syste
c 34	29	0.9	1139	22	AAK86252	Human immune/haema
c 35	29	0.9	1873	21	AACT76766	Human ORFX ORF2321
c 36	29	0.9	2586	22	ABAI4588	Human nervous syst
c 37	29	0.9	3270	22	AAS26707	Human genomic DNA
c 38	29	0.9	4978	22	AAK89129	Human digestive sy
c 39	29	0.9	4978	22	AAS31879	Human liver associ
c 40	29	0.9	5690	22	AAS45368	Chemically pretrea
c 41	29	0.9	5752	22	AAF85087	Nucleotide sequenc
c 42	29	0.9	6025	22	AAS45338	Chemically pretrea
c 43	29	0.9	6025	22	AAS45339	Chemically pretrea
c 44	29	0.9	6250	22	AAL36600	Human musculoskele
c 45	29	0.9	6251	22	AAL36603	Human musculoskele

ALIGNMENTS

RESULT 1
AAT34620
ID AAT34620 standard; DNA; 3337 BP.
XX
AC AAT34620;
XX
DT 12-NOV-1996 (first entry)
XX
DE P. vivax ESP-1 blood stage antigen coding sequence.
XX
KW ESP-1; blood stage antigen; diagnosis; malaria; infection;
KW causative agent; antibody; monoclonal; polyclonal; assay; ds.
XX
OS Plasmodium vivax (clone PvMB3.3.1).
XX
FH Key Location/Qualifiers
FT Exon 1..91
FT /*tag= a
FT /*note= "encodes initial (N-terminal) sequence of
FT intron 92..230 hydrophobic amino acids"
FT /*tag= b
FT exon 231..3197
FT /*tag= c
XX
XX US5532133-A.
XX
PD 02-JUL-1996.
XX
PF 02-JUN-1993; 93US-0072610.

Qy	1741	agatccgtatgcagagagggctgtataacagttaccatcaaaaggaagccctgtacaagtac	1800
Db	1741	agatccgtatgcagagagggctgtataacagttaccatcaaaaggaagccctgtacaagtac	1800
Qy	1801	caatggcagttagggccgcgaagaagtgccaacggaagaatttatcaactccaagagg	1860
Db	1801	caatggcagttagggccgcgaagaagtgccaacggaagaatttatcaactccaagagg	1860
Qy	1861	acgatcttcaatttagaaggaactgcagaagctccagaggaagggagaatttagtattagaag	1920
Db	1861	acgatcttcaatttagaaggaactgcagaagctccagaggaagggagaatttagtattagaag	1920
Qy	1921	gagaaggaagaccacgcgaagaagcgcaagaaggaaggagaccacagaagggagaagtgc	1980
Db	1921	gagaaggaagaccacgcgaagaagcgcaagaaggaaggagaccacagaagggagaagtgc	1980
Qy	1981	cagaagaagaatttagaggaactccagaggaacttcgcgaatttagaagaaccacagagg	2040
Db	1981	cagaagaagaatttagaggaactccagaggaacttcgcgaatttagaagaaccacagagg	2040
Qy	2041	aagaagtagaagaacccgtagagggcggaagaanaactgcgaagaaggagaagaagtggaaagg	2100
Db	2041	aagaagtagaagaacccgtagagggcggaagaanaactgcgaagaaggagaagaagtggaaagg	2100
Qy	2101	tacctgcagaagttagaagaagtggaaagggtacctgcgagaagttagaagaagtggaagagg	2160
Db	2101	tacctgcagaagttagaagaagtggaaagggtacctgcgagaagttagaagaagtggaagagg	2160
Qy	2161	taccagaagaagttagaagaaggttaccgcgagaagttagaagaagtggaagaaggtaccagaag	2220
Db	2161	taccagaagaagttagaagaaggttaccgcgagaagttagaagaagtggaagaaggtaccagaag	2220
Qy	2221	aagtggaaaggggtaccagaagaagtggaagaaggttaccagaagaagtggaagaaggtaccag	2280
Db	2221	aagtggaaaggggtaccagaagaagtggaagaaggttaccagaagaagtggaagaaggtaccag	2280
Qy	2281	aagaagtggaaaggttggaagaagttagaagaagtagaggttaccagcggttagtagaagtag	2340
Db	2281	aagaagtggaaaggttggaagaagttagaagaagtagaggttaccagcggttagtagaagtag	2340
Qy	2341	aagtaccagcggttagtagaagaagaggttcgcagaagaagttagaagaagaagaagaagagg	2400
Db	2341	aagtaccagcggttagtagaagaagaggttcgcagaagaagttagaagaagaagaagaagagg	2400
Qy	2401	aagaaccagttagaggaagagatgtattacaatttagttaaccatcggaagaagatatatc	2460
Db	2401	aagaaccagttagaggaagagatgtattacaatttagttaaccatcggaagaagatatatc	2460
Qy	2461	aattagacaaccagaagaagcgaatttaggctctgggaattttatctatcatcgacatgc	2520
Db	2461	aattagacaaccagaagaagcgaatttaggctctgggaattttatctatcatcgacatgc	2520
Qy	2521	actaccaagacgttccaaaggaatttatggaaagaagaagaacgtcgagtgatccat	2580
Db	2521	actaccaagacgttccaaaggaatttatggaaagaagaagaacgtcgagtgatccat	2580
Qy	2581	tgaaccagaagaatttgcgaaggaagattcacaacttcagaagaatggtctcaacttcatc	2640
Db	2581	tgaaccagaagaatttgcgaaggaagattcacaacttcagaagaatggtctcaacttcatc	2640
Qy	2641	aaggcttagaagcgactgggaacgattagaagtgagcttaataaaggcttagagaagaagt	2700
Db	2641	aaggcttagaagcgactgggaacgattagaagtgagcttaataaaggcttagagaagaagt	2700
Qy	2701	ggatggaaacaaagaataaagaattgggctgggctgggcttcgcttaattgaaaaataatcggt	2760
Db	2701	ggatggaaacaaagaataaagaattgggctgggctgggcttcgcttaattgaaaaataatcggt	2760
Qy	2761	cagaaataatgtccaaatttcacaaagaagaagccccagctgggttttagaaaaacagagat	2820
Db	2761	cagaaataatgtccaaatttcacaaagaagaagccccagctgggttttagaaaaacagagat	2820

RESULT 2

RESOL
AAX15174

ID AAX15174 standard; DNA; 3337 BP.

XX AAX15174:

AC
XX
AAAL31747

DT 28-APR-1999 (first entry)

XX

DE DNA encoding a secreted blood-stage protein called PvESP-1.

XX XX

KW Erythrocyte secreted protein-1; PvESP-1; malarial antigen;

KW blood-stage prote

20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854

OS Plasmodium vivax.
VV

XX	Key	Location/Qualifiers
FH		

FT	key	LOCATION
FT	CDS	1...3197

```
ET
CDS
I...3137
/*tag=
```

FT /note= "contains 1 intron"

FT	exon	1..91
FT	exon	1..91

ET /*tag=

ET	/number
1	1
2	2
3	3
4	4
5	5
6	6
7	7
8	8
9	9
10	10
11	11
12	12
13	13
14	14
15	15
16	16
17	17
18	18
19	19
20	20
21	21
22	22
23	23
24	24
25	25
26	26
27	27
28	28
29	29
30	30
31	31
32	32
33	33
34	34
35	35
36	36
37	37
38	38
39	39
40	40
41	41
42	42
43	43
44	44
45	45
46	46
47	47
48	48
49	49
50	50
51	51
52	52
53	53
54	54
55	55
56	56
57	57
58	58
59	59
60	60
61	61
62	62
63	63
64	64
65	65
66	66
67	67
68	68
69	69
70	70
71	71
72	72
73	73
74	74
75	75
76	76
77	77
78	78
79	79
80	80
81	81
82	82
83	83
84	84
85	85
86	86
87	87
88	88
89	89
90	90
91	91
92	92
93	93
94	94
95	95
96	96
97	97
98	98
99	99
100	100

FT	Intron	92..230

FT 2 /*tag=

FT	number
FT	21

```
FT      exon
231...31
/*tag=
```

ET. /number
ET. /tag=

XX
FI
XX
Tadumil /

PN US5874527-A.

XX

PD 23-FEB-1999.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PF 30-SEP-1996;

XX
DP 02-JUN-1963.

PR 02-JUN-1993;

PR 07-JUN-1995; 95US-0478417.
 PR 30-SEP-1996; 96US-0719822.
 XX
 PA (UYNV) UNIV NEW YORK STATE.
 XX
 PI Barnwell JW;
 XX
 DR WPI; 1999-180063/15.
 DR P-PSDB; AAW97039.
 XX
 XX Plasmodium vivax peptide antigen - for diagnosis of malaria caused
 PT by Plasmodium vivax
 XX
 PS Example 4; Fig 5A-C; 23pp; English.
 XX
 CC The present sequence encodes a C-terminal erythrocyte secreted
 CC protein-1 (PvESP-1) of Plasmodium vivax. PvESP-1 is a malarial
 CC antigen which is a secreted blood-stage protein present in detectable
 CC amounts in biological samples from individuals infected with P. vivax.
 CC The protein comprises an epitope not present in other Plasmodium species
 CC that cause malaria in humans, and is bound by monoclonal antibody
 CC 1D1G10. The peptide antigen can be used in immunoassays for diagnosis
 CC of malaria caused by P. vivax and/or can be used to produce antibodies
 CC for use in such immunoassays.
 XX
 SQ Sequence 3337 BP; 1304 A; 467 C; 875 G; 691 T; 0 other;

Query Match 100.0%; Score 3337; DB 20; Length 3337;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	gaattccggtaaagttaacactatggttcgtatctatataataacacttaatttacc	60
DB	1	gaattccggtaaagttaacactatggttcgtatctatataataacacttaatttacc	60
QY	61	ttttgttttttttaattcatgcttcaacagtaacagtaataataataataataaactgc	120
DB	61	ttttgttttttttaattcatgcttcaacagtaacagtaataataataataataaactgc	120
QY	121	tatatatacatatattcataagtgccattgtggaattgcgatactattaaatttacgta	180
DB	121	tatatatacatatattcataagtgccattgtggaattgcgatactattaaatttacgta	180
QY	181	aaacaataattgaaaaaatt	240
DB	181	aaacaataattgaaaaaatt	240
QY	241	aattgaaaaatgcttctgatgattgtgtgaggtgagagatccttcaaacgacggtttag	300
DB	241	aattgaaaaatgcttctgatgattgtgtgaggtgagagatccttcaaacgacggtttag	300
QY	301	aattagaagaggaaaattttgatgagaattcaggtgatgataaactcttttagatgcta	360
DB	301	aattagaagaggaaaattttgatgagaattcaggtgatgataaactcttttagatgcta	360
QY	361	ccccgaagatgactttgcttcaacagatttgccaattgaacagcgtatgaggaaagtcacg	420
DB	361	ccccgaagatgactttgcttcaacagatttgccaattgaacagcgtatgaggaaagtcacg	420
QY	421	aaacgttagatgaggtgaatcatattgagagaggtttccactgaagatatggaacagaag	480
DB	421	aaacgttagatgaggtgaatcatattgagagaggtttccactgaagatatggaacagaag	480
QY	481	atggctcaacagatgatacggaacagaagaagactacgttgatgataaggaggagaag	540
DB	481	atggctcaacagatgatacggaacagaagaagactacgttgatgataaggaggagaag	540
QY	541	aagaagctggcgatatggaagcaggggaagaagctggtgatttggaaagcaggggaagaa	600
DB	541	aagaagctggcgatatggaagcaggggaagaagctggtgatttggaaagcaggggaagaa	600
QY	601	ctggcatttggaaagcaggggaagaactggcgatttggaaagcaggggaagaactggtg	660

Db 1681 ataacgatttaacgtatgcagacattacatcctttgacccattatttaacaaatcctca 1740
Qy 1741 aggatccctgatacgaagagagctgtacacagtaccatcaaaaggaagcactgtacaaagtac 1800
Db 1741 aggatccctgatacgaagagagctgtacacagtaccatcaaaaggaagcactgtacaaagtac 1800
Qy 1801 cagtggcagtagggcccgcaagaagtgcacaacggaagaattgtagcaactccaagagg 1860
Db 1801 cagtggcagtagggcccgcaagaagtgcacaacggaagaattgtagcaactccaagagg 1860
Qy 1861 acgatttcgaattagaagaactgcagaagctccagaagaaggagaattagattagaga 1920
Db 1861 acgatttcgaattagaagaactgcagaagctccagaagaaggagaattagattagaga 1920
Qy 1921 gagaaggaacccaacggaagaagccaagaagaggaagaggaagccaagaggaagtgcc 1980
Db 1921 gagaaggaacccaacggaagaagccaagaagaggaagaggaagccaagaggaagtgcc 1980
Qy 1981 cagaagaagaattagaggcaactccagaggacgatttcgaattagaagaaccaaagagag 2040
Db 1981 cagaagaagaattagaggcaactccagaggacgatttcgaattagaagaaccaaagagag 2040
Qy 2041 aagaagtagaagaacccgtagaggcggaagaactgcagaaggagaagaagtggaaagg 2100
Db 2041 aagaagtagaagaacccgtagaggcggaagaactgcagaaggagaagaagtggaaagg 2100
Qy 2101 tacctgcagaagtagaagaagtggaaagggtaccctgcagaagtagaagaagtggaaagg 2160
Db 2101 tacctgcagaagtagaagaagtggaaagggtaccctgcagaagtagaagaagtggaaagg 2160
Qy 2161 taccagaagaagtagaagaagtcaccgcagaagtagaagaagtggaagaggtaccagaag 2220
Db 2161 taccagaagaagtagaagaagtcaccgcagaagtagaagaagtggaagaggtaccagaag 2220
Qy 2221 aagtggaaagaggtaccagaagaagtggaaagggtaccagaagaagtggaagaggtaccag 2280
Db 2221 aagtggaaagaggtaccagaagaagtggaaagggtaccagaagaagtggaagaggtaccag 2280
Qy 2281 aagaagtggaagaagtggaagaagtagaagaagtagagaggtaccagcgttagtagaagtag 2340
Db 2281 aagaagtggaagaagtggaagaagtagaagaagtagagaggtaccagcgttagtagaagtag 2340
Qy 2341 aagtaccagcgttagtagaagaagaaggtgccagaagaagtagaagaagaagaagaagagg 2400
Db 2341 aagtaccagcgttagtagaagaagaaggtgccagaagaagtagaagaagaagaagaagagg 2400
Qy 2401 aagaaccagtagagaagaagaagtattacaattagtaataaccatcggaagaagatatac 2460
Db 2401 aagaaccagtagagaagaagaagtattacaattagtaataaccatcggaagaagatatac 2460
Qy 2461 aattagacaaaccaaagaagaacgaattaggctctgggaattttatctatcatcgacatgc 2520
Db 2461 aattagacaaaccaaagaagaacgaattaggctctgggaattttatctatcatcgacatgc 2520
Qy 2521 actaccagaagcttccaagaagaatttagaagaagaagaagaagaactcagtgatccat 2580
Db 2521 actaccagaagcttccaagaagaatttagaagaagaagaagaagaactcagtgatccat 2580
Qy 2581 tgaaccagaagaattttccaagaagaattcacatctcacagaatgctcacattcattc 2640
Db 2581 tgaaccagaagaattttccaagaagaattcacatctcacagaatgctcacattcattc 2640
Qy 2641 aaggcctgaagcagctgggaacgatttagaagtagagcttaataaggctagagagaagat 2700
Db 2641 aaggcctgaagcagctgggaacgatttagaagtagagcttaataaggctagagagaagat 2700
Qy 2701 ggaatggaacaaagaataaaagaatggctggctggctcgtcttaattgaaaaataaattggt 2760
Db 2701 ggaatggaacaaagaataaaagaatggctggctggctcgtcttaattgaaaaataaattggt 2760
Qy 2761 cagaatatagtcataatttcaacaaagaagaagaccagctgggtttgagaacagagagt 2820
Db 2761 cagaatatagtcataatttcaacaaagaagaagaccagctgggtttgagaacagagagt 2820

Qy 2821 ggagcgacgagaaatggaaaaatggtttaaagcagaagtcaaatcccaaatgtattcac 2880
Db 2821 ggagcgacgagaaatggaaaaatggtttaaagcagaagtcaaatcccaaatgtattcac 2880
Qy 2881 acttgaaaaatggatgaacgacactcattccaaattttataaaattcttctgaaagata 2940
Db 2881 acttgaaaaatggatgaacgacactcattccaaattttataaaattcttctgaaagata 2940
Qy 2941 tgtcacaaatttgaaacaaagaacaaacaaagaatggttaatactcactggaaaaagaacg 3000
Db 2941 tgtcacaaatttgaaacaaagaacaaacaaagaatggttaatactcactggaaaaagaacg 3000
Qy 3001 aacgggggttatggttctgaatcatttggaagttatgaccacatcaaaattattaaatgtgg 3060
Db 3001 aacgggggttatggttctgaatcatttggaagttatgaccacatcaaaattattaaatgtgg 3060
Qy 3061 ctaagagtcgagaatgggtaccgtgccatcctaaataataatagagaagaagagaactca 3120
Db 3061 ctaagagtcgagaatgggtaccgtgccatcctaaataataatagagaagaagagaactca 3120
Qy 3121 tgaatgggtttctcctaagaacaaacgaatatttaggcaacaaagaatggaataatggact 3180
Db 3121 tgaatgggtttctcctaagaacaaacgaatatttaggcaacaaagaatggaataatggact 3180
Qy 3181 cattggaaaaaagtttaaatttttggttcattcaattcgaatcgtcacacattttcttgaaaa 3240
Db 3181 cattggaaaaaagtttaaatttttggttcattcaattcgaatcgtcacacattttcttgaaaa 3240
Qy 3241 cgcttaaccaaaggagaatggaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaat 3300
Db 3241 cgcttaaccaaaggagaatggaatcgaatcgaatcgaatcgaatcgaatcgaatcgaatcgaat 3300
Qy 3301 aagaagaacagattattctcttataaaataaataatc 3337
Db 3301 aagaagaacagattattctcttataaaataaataatc 3337

RESULT 3
AAH76457
ID AAH76457 standard; DNA; 3337 BP.
XX
AC AAH76457;
XX
DT 22-OCT-2001 (first entry)
XX
DE Plasmodium vivax ESP-1 DNA.
XX
KW Plasmodium vivax; ESP-1; erythrocyte secreted protein-1; PvESP-1;
KW species-specific; malarial peptide antigen; infection; diagnosis;
KW malaria; ds.
XX
OS Plasmodium vivax.
XX
FH Key Location/Qualifiers
FT CDS 1..3197
FT FT /*tag= a
FT FT /*product= "ESP-1"
FT FT exon 1..91
FT FT /*tag= b
FT FT /*number= 1
FT FT 92..230
FT FT intron
FT FT /*tag= c
FT FT /*number= 1
FT FT 231..3197
FT FT /*tag= d
FT FT /*number= 2
XX
PN US6231861-B1.
XX
PD 15-MAY-2001.
XX
PP 05-JUN-1998; 98US-0092458.


```
RESULT 11
ABL09674/c
ID ABL09674 standard; cDNA; 5856 BP.
XX
AC ABL09674;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster expressed polynucleotide SEQ ID NO 23504.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ss.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
XX
XX 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX
XX WPI; 2001-658660/75.
XX
XX P-PSDB; ABB65571.
XX
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
XX genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX
XX Claim 1; SEQ ID NO 23504; 2lpp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
XX capable of detecting 1000 or more genes from Drosophila. The invention is
XX useful in developmental biology and in elucidating cell signalling and
XX cell-cell interactions in higher eukaryotes for the development of
XX insecticides, therapeutics and pharmaceutical drugs. The invention
XX discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
XX sequences (ABL01840-ABL16175) and the encoded proteins
XX (ABB57737-ABB72072).
XX
XX The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 5856 BP; 1527 A; 1429 C; 1445 G; 1455 T; 0 other;

Query Match 0.9%; Score 30; DB 23; Length 5856;
Best Local Similarity 100.0%; Pred. No. 0.076;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 195 aaaaatttttttttttttttttttttttttttttt 224
    | | | | | | | | | | | | | | | | | | | |
Db 3372 AAAAAATTTTTTTTTTTTTTTTTTTTGTGT 3343

RESULT 12
AAS45437
ID AAS45437 standard; DNA; 8801 BP.
XX
AC AAS45437;
XX
DT 18-DEC-2001 (first entry)
XX
DE Chemically pretreated complementary DNA associated with cell cycle #71.
XX
XX Cell cycle; human; CpG dinucleotide; cytosine methylation; HIV; aging;
KW
```

```
KW human immunodeficiency virus; neurodegenerative disorder; solid tumour;
KW graft-versus-host disease; glomerular disease; Lewy body disease; cancer;
KW arthritis; arteriosclerosis; anti-HIV; neuroprotective; antiarthritic;
KW immunosuppressive; antitumour; cytostatic; antiarteriosclerotic; ds;
KW PCR primer.
XX
XX Homo sapiens.
XX
XX WO200168911-A2.
XX
XX 20-SEP-2001.
XX
XX 15-MAR-2001; 2001WO-EP02945.
XX
XX 15-MAR-2000; 2000DE-1013847.
XX
XX 06-APR-2000; 2000DE-1019058.
XX
XX 07-APR-2000; 2000DE-1019173.
XX
XX 30-JUN-2000; 2000DE-1032529.
XX
XX 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K;
XX
XX WPI; 2001-602751/68.
XX
XX Designing primers and probes for analysing diseases associated with
XX cytosine methylation state e.g. arthritis, cancer, aging,
XX arteriosclerosis comprising fragments of chemically modified genes
XX associated with cell cycle -
XX
XX Claim 1; SEQ ID No 142; 28pp; English.
XX
XX Sequences AAS45296-AAS45520 represent chemically pretreated genomic DNA
XX molecules associated with the cell cycle and specific PCR primers of the
XX invention. The sequences are useful for detecting the methylation state
XX of all CpG dinucleotides in a sequence and therefore for analysing
XX associated diseases. By analysing cytosine methylations in the pretreated
XX DNA, genetic and/or epigenetic parameters for the diagnosis and therapy
XX of existing diseases or the predisposition to specific diseases can be
XX ascertained. The parameters may be compared to another set of genetic
XX and/or epigenetic parameters, the differences serving as basis for
XX diagnosis and/or prognosis events which are disadvantageous to patients.
XX The sequences of the invention are useful for the diagnosis and therapy
XX of HIV infection, neurodegenerative disorders, graft-versus-host disease,
XX aging, glomerular disease, Lewy body disease, arthritis,
XX arteriosclerosis, solid tumours and cancers.
XX
XX Sequence 8801 BP; 2344 A; 366 C; 2057 G; 4034 T; 0 other;

Query Match 0.9%; Score 30; DB 22; Length 8801;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 192 gaaaaaatttttttttttttttttttttttttttttt 221
    | | | | | | | | | | | | | | | | | | | |
Db 971 gaaaaaatttttttttttttttttttttttttttttt 1000

RESULT 13
ABL33741
ID ABL33741 standard; DNA; 8801 BP.
XX
XX ABL33741;
XX
XX 26-MAR-2002 (first entry)
XX
XX Human immune system associated gene SEQ ID NO: 1714.
XX
XX Human; immune system disease; cytosine methylation; antiasthmatic;
XX antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
```

KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
KW gene; ds.

XX Homo sapiens.

OS WO200200928-A2.

PN 03-JAN-2002.

PD 02-JUL-2001; 2001WO-EP07537.

PF 30-JUN-2000; 2000DE-1032529.

PR 01-SEP-2000; 2000DE-1043826.

XX (EPIG-) EPIGENOMICS AG.

PA Olek A, Piepenbrock C., Berlin K;

PI WPI; 2002-130909/17.

DR Nucleic acid comprising fragment of chemically modified gene, useful

PT for diagnosis and treatment of diseases associated with abnormal

PT cytosine methylation

XX Claim 1; SEQ ID NO 1714; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated

CC genes which are modified by the methylation of cytosines. The sequences

CC can be used in the diagnosis and treatment of immune system disorders,

CC including eye diseases such as retinopathy, neovascular glaucoma and

CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid

CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,

CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel

CC diseases. The present sequence is a gene of the invention.

XX Sequence 8801 BP; 2344 A; 366 C; 2057 G; 4034 T; 0 other;

SQ Query Match 0.9%; Score 30; DB 24; Length 8801;

Best Local Similarity 100.0%; Pred. No. 0.072;

Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 192 gaaaaaattttttttttttttttttttttttt 221

|||||

Db 971 gaaaaaattttttttttttttttttttttttt 1000

RESULT 14

AAQ43973

ID AAQ43973 standard; DNA; 32 BP.

XX AC AAQ43973;

XX 28-OCT-1993 (first entry)

DE Triple helix forming oligonucleotide I.

XX Purine; pyrimidine; tracts; intramolecular triplex; therapeutic;

KW diagnostic; control; gene expression; mRNA synthesis suppression;

XX ss.

XX Synthetic.

XX WO9312230-A.

PN 24-JUN-1993.

PR 21-JAN-1992; 92US-0826934.
XX (STRI) SRI INT.
PA Jayasena SD, Johnston BH;
XX WPI; 1993-214172/26.
DR New oligo:nucleotide(s) forming triple helix with target nucleic
XX acid - contain purine and pyrimidine tracts in specific
PT orientations, useful therapeutically or diagnostically e.g. for
PT inactivating HIV RNA, etc.
XX Disclosure; Page 47; 101pp; English.
XX The sequence is that of an oligonucleotide, I, which is able to form
CC a triple helix with a duplex nucleic acid (dsNA) contg. a target
CC sequence which comprises at least one pyrimidine tract, and at least
CC one adjacent purine tract. It is useful for therapeutic or
CC diagnostic control of gene expression, e.g. suppression of mRNA
CC synthesis from a target gene. A specified application is targeting
CC of RNA in the HIV-1 genome. When appropriately labelled it may also
CC be used as a probe. Attachment of cleavage agents caused permanent
CC inactivation of the target by site-specific cleavage.
XX Sequence 32 BP; 8 A; 0 C; 0 G; 24 T; 0 other;
SQ Query Match 0.9%; Score 29; DB 14; Length 32;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 193 aaaaaattttttttttttttttttttttttt 221
|||||
Db 2 aaaaaattttttttttttttttttttttttt 30
RESULT 15
AAQ25032
ID AAQ25032 standard; DNA; 40 BP.
XX AC AAQ25032;
XX 13-JUL-1992 (first entry)
DT Oligonucleotide specific for HIV proviral DNA.
DE HIV; thiolation; reverse transcriptase; primer; inhibition;
XX homooligomer; ss.
XX Synthetic.
XX WO9203127-A.
XX 05-MAR-1992.
XX 15-AUG-1991; 91WO-US05919.
XX 16-AUG-1990; 90US-0568131.
XX (UWNY-) RES FOUND UNIV NEW.
XX Bardos TJ, Ho YK, Aradi J, Schinazi RF;
PI WPI; 1992-096567/12.
DR Compsn. contg. 5-thiolated (oligo-poly-)-nucleotide(s) - for
XX treating HIV infection, AIDS and for preventing HIV-1 infection.
XX Disclosure; Page 11; 42pp; English.
XX The oligomer comprises a non-thiolated (binding) homooligonucleotide
CC region (d(A)12) to promote the binding of the remaining portion of

QY 193 aaaaaattttttttttttttt 221

```
Db 337 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 309

RESULT 18
AAI85033/c
ID AAI85033 standard; cDNA; 344 BP.
XX
AC AAI85033;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5093.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR P-PSDB; AAO05102.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 5093; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 344 BP; 165 A; 55 C; 74 G; 50 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaatttttttttttttttttttttttttttttt 221
DB 158 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTT 130

RESULT 19
AAI91956/c
ID AAI91956 standard; cDNA; 352 BP.
XX

Db 337 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 309

RESULT 18
AAI85033/c
ID AAI85033 standard; cDNA; 344 BP.
XX
AC AAI85033;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 5093.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR P-PSDB; AAO05102.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 5093; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 344 BP; 165 A; 55 C; 74 G; 50 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 344;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaatttttttttttttttttttttttttttttt 221
DB 158 AAAAAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTT 130

RESULT 19
AAI91956/c
ID AAI91956 standard; cDNA; 352 BP.
XX

AC AAI91956;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 12016.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX
OS Homo sapiens.
XX
PN WO200164835-A2.
XX
PD 07-SEP-2001.
XX
PF 26-FEB-2001; 2001WO-US04927.
XX
PR 28-FEB-2000; 2000US-0515126.
XX
PR 18-MAY-2000; 2000US-0577409.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-514838/56.
XX
DR P-PSDB; AAO12025.
XX
PT Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
PS Claim 1; SEQ ID NO 12016; 1399pp + Sequence Listing; English.
XX
CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 352 BP; 132 A; 61 C; 66 G; 93 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 352;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 193 aaaaaatttttttttttttttttttttttttttttt 221
DB 89 AAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 61

RESULT 20
AAI84186/c
ID AAI84186 standard; cDNA; 361 BP.
XX
AC AAI84186;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 4246.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
```

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO04255.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX Claim 1; SEQ ID NO 4246; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 361 BP; 110 A; 102 C; 61 G; 83 T; 5 other;
SQ
Query Match 0.9%; Score 29; DB 22; Length 361;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 193 aaaaaatttttttttttttttttttttttttttttt 221
Db 216 AAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 188
RESULT 21
AAI84071/c
ID AAI84071 standard; cDNA; 369 BP.
XX AC AAI84071;
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 4131.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200164835-A2.

XX 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-514838/56.
XX P-PSDB; AAO04140.
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX Claim 1; SEQ ID NO 4131; 1399pp + Sequence Listing; English.
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating t
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 369 BP; 139 A; 56 C; 83 G; 90 T; 1 other;
SQ
Query Match 0.9%; Score 29; DB 22; Length 369;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 193 aaaaaatttttttttttttttttttttttttttttt 221
Db 276 AAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 248
RESULT 22
AAI82188/c
ID AAI82188 standard; cDNA; 404 BP.
XX AC AAI82188;
XX 06-NOV-2001 (first entry)
XX Human polynucleotide SEQ ID NO 2248.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200164835-A2.
XX PD 07-SEP-2001.
XX PF 26-FEB-2001; 2001WO-US04927.
XX PR 28-FEB-2000; 2000US-0515126.
XX PR 18-MAY-2000; 2000US-0577409.

```
XX (HYSE-) HYSEQ INC.
XX PA
XX PI Tang YT, Liu C, Drmanac RT;
XX DR WPI; 2001-514838/56.
XX DR P-PSDB; AAO02257.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 2248; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 404 BP; 133 A; 74 C; 83 G; 114 T; 0 other;
XX
XX
XX Query Match 0.9%; Score 29; DB 22; Length 404;
XX Best Local Similarity 100.0%; Pred. No. 0.25;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 194 aaaaaattttttttttttttttttttttttt 222
XX ||||||||||||||||||||||||||||
XX Db 195 AAAAAATTTTttttttttttttttttttttt 167
XX
XX
XX RESULT 23
XX AAI87091/c
XX ID AAI87091 standard; cDNA; 406 BP.
XX
XX AC AAI87091;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 7151.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200164835-A2.
XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US04927.
XX
XX PR 28-FEB-2000; 2000US-0515126.
XX
XX PR 18-MAY-2000; 2000US-0577409.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AAO07160.
XX
```

```
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 7151; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
XX the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 406 BP; 127 A; 115 C; 66 G; 98 T; 0 other;
XX
XX
XX Query Match 0.9%; Score 29; DB 22; Length 406;
XX Best Local Similarity 100.0%; Pred. No. 0.25;
XX Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 193 aaaaaattttttttttttttttttttttttt 221
XX ||||||||||||||||||||||||
XX Db 287 AAAAAATTTTttttttttttttttttttttt 259
XX
XX
XX RESULT 24
XX AAI87993/c
XX ID AAI87993 standard; cDNA; 416 BP.
XX
XX AC AAI87993;
XX
XX DT 06-NOV-2001 (first entry)
XX
XX DE Human polynucleotide SEQ ID NO 8053.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorders; arthritis; inflammation; ss.
XX
XX OS Homo sapiens.
XX
XX PN WO200164835-A2.
XX
XX PD 07-SEP-2001.
XX
XX PF 26-FEB-2001; 2001WO-US04927.
XX
XX PR 28-FEB-2000; 2000US-0515126.
XX
XX PR 18-MAY-2000; 2000US-0577409.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX PI Tang YT, Liu C, Drmanac RT;
XX
XX WPI; 2001-514838/56.
XX
XX P-PSDB; AAO08062.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX diagnosing and treating e.g. leukaemia, inflammation and immune
XX disorders -
XX
XX Claim 1; SEQ ID NO 8053; 1399pp + Sequence Listing; English.
XX
```

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO3910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
xx
SQ Sequence 416 BP: 208 A; 33 C; 96 G; 62 T; 17 other:
xx

```
Query Match          0.9%; Score 29; DB 22; Length 416;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	193	aaaaaaatttttttttttttttttttttttt 221 	221
Db	271	AAAAAAATTTTTTTTTTTTTTTTTTTTTTTT 243	243

RESULT 25
AAI84767/C
ID AAI84767 standard; cDNA; 417 BP.
XX
XX
XX AAI84767;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human polynucleotide SEQ ID NO 4827.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation; ss.

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_sequences.
XX
XX Sequence 417 BP: 149 A; 80 C; 95 G; 93 T; 0 other;

Query Match 0.9%; Score 29; DB 22; Length 417;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 29; Conservative 0; Mismatches 0; Indels

[illegible]

RESULT 26
AAI89825/C
ID AAI89825 standard; CDNA: 422 BP.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 15:37:31 ; Search time 4065.53 Seconds
(without alignments)
17176.568 Million cell updates/sec

Title: US-09-667-130-1
Perfect score: 3337
Sequence: 1 gattccgtaagtaacaa.....tcttataaaataaataattc 3337

Scoring table: OLIGO_NUC
Gapop 60.0 , Capext 60.0

Searched: 1797656 seqs, 10463268293 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htgo_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
------------	-------------	-------	--------	----	-------------

1	3337	100.0	3337	6	ARI51085
2	3337	100.0	3337	6	I23337
3	40	1.2	70206	2	AC067910
4	40	1.2	165868	2	AC027625
5	34	1.0	186562	2	AL451079
6	34	1.0	200898	2	AC096541
7	34	1.0	225609	2	AL645546
8	33	1.0	6862	6	AX251053
9	33	1.0	6862	6	AX251776
10	33	1.0	6862	6	AX345124
11	33	1.0	240000	9	HS421735
12	33	1.0	246176	2	AC096301
13	32	1.0	224010	2	AP001848
14	32	1.0	318221	2	PFMA113P3
15	31	0.9	35836	3	CEZC334
16	31	0.9	140798	2	AC027803
17	31	0.9	172896	9	AL158193
18	31	0.9	173817	2	AC012141
19	31	0.9	198820	2	AC023315
20	31	0.9	198932	2	AL671090
21	31	0.9	243675	10	AC079869
22	30	0.9	8801	6	AX277981
23	30	0.9	8801	6	AX323672
24	30	0.9	8801	6	AX346643
25	30	0.9	16689	9	HS4011712
26	30	0.9	49816	9	AC093149
27	30	0.9	58569	2	AC015033
28	30	0.9	64255	2	AC024343
29	30	0.9	67462	2	AC091595
30	30	0.9	73858	2	AC109316
31	30	0.9	82419	9	AC004979
32	30	0.9	89203	2	AC010691
33	30	0.9	92491	2	AL672088
34	30	0.9	96874	9	AL138709
35	30	0.9	96879	9	HS384F21
36	30	0.9	98935	9	AC004935
37	30	0.9	99187	9	HS35M08
38	30	0.9	103574	9	HSAC002115
39	30	0.9	113880	3	PFMA13P4
40	30	0.9	115027	2	AP001480
41	30	0.9	117309	2	AC099091
42	30	0.9	123189	9	AC002556
43	30	0.9	125722	2	AC015695
44	30	0.9	138800	30	AC016215
45	30	0.9	140955	2	AP001971

ALIGNMENTS

RESULT	1	ARI51085	Sequence 1 from patent US 6231861.	3337 bp	DNA	linear	PAT 08-AUG-2001
LOCUS	ARI51085	Sequence 1 from patent US 6231861.					
DEFINITION	ARI51085	Sequence 1 from patent US 6231861.					
ACCESSION	ARI51085	Sequence 1 from patent US 6231861.					
VERSION	ARI51085.1	GI:15117135					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 3337)						
AUTHORS	Barnwell,J.W.						
TITLE	Plasmodium vivax blood stage antigens, antibodies, and diagnostic assays						
JOURNAL	Patent: US 6231861-A 1 15-MAY-2001;						
FEATURES	Location/Qualifiers						
source	1. 3337						
BASE COUNT	1304 a 467 c 875 g 691 t						
ORIGIN							

Query Match 100.0%; Score 3337; DB 6; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;		Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	1	gaattccggtaaaagtaacaactatagtttctgatatatatataaacttactaattttatc	60
Db	1	GAATTCGGGTAAAGTAACAACATATGTTTCGTATCTATATATACTTACTAAATTTTATC	60
Qy	61	ttttgctttctttaaattcattcgtctcaacagtaagataaaaaataatctataaaaactgc	120
Db	61	TTTTGCTTTCTTTAAATTCATGCTTCAACAGTAGATATAAAATTAATCTATAAAAACATGC	120
Qy	121	tatatatacatatattcataaagtggttggaatttggaattgcgatcatttaatttacctata	180
Db	121	TATATATACATATATATTCATAAAGTGGCATTTGTGAATTTGGCATCATTTAAATTTACGTA	180
Qy	181	aaaaacaattatgaaaaaattttttttttttttttttttttttttttttttgtctacagaacgttag	240
Db	181	AAAACAATATTGAAAAAATTTTTTFTTTTFTTTTFTTTTGTCTACAGAACGATTTAG	240
Qy	241	aattggaaaatgcttctgatgatgttgtagaggtgagagatccttcaaacgacggttttag	300
Db	241	AATTGGAAAAATGCTTCTGATGATGTTGTAGAGGTGAGGATCCCTTCAAACGACGCTTTAG	300
Qy	301	aattagaagagaaaattttgatgagaattcagggtgatgatagaactcttttagatgcta	360
Db	301	AATTAGAAGAGGAAAATTTTGATGAGAATTCAGGTGATGATGAATACTCTTTTAGATGCTA	360
Qy	361	ccccgaagatgaactttccttaacagatttgcgaattgaagacgatgagaaagtcacag	420
Db	361	CCCCCGAAGATGACATTTGCCTTAACAGATTTGCCAANTTGAAGACGATGAGGAAGTCAAG	420
Qy	421	aaacttagatgaggtgaattcattagagaggtttccactgaagatgatgaaacagaag	480
Db	421	AAACCTTAGATGGAGTGAATCATTTAGGAGAGCTTTTCCACTGAACATATGGAACAGAAG	480
Qy	481	atggctcaacagatgatacaggaacacagaagaagactacctggtgatgatgaaaggagaag	540
Db	481	ATGGCTCAACAGATGATACGGAACAGAGAAAGAGGACTACCTGGTGATATGGAAGGAGAAG	540
Qy	541	aagaagctggcgataggaagcaggggaagaagcttggtgattggaagcaggggaagaaa	600
Db	541	AAGAAGCTGGCGATATGGAAGCAGGGGAAGAAGCTGGTGATTGGAACAGCGGGAAGAA	600
Qy	601	ctggcgatttgaagcaggggaagaactggcgatttggaaagcaggggaagaagctggtg	660
Db	601	CTGGCGATTTTGGAACGAGGGGAAGAACTGGCGATTTTGGAAGCAGGGGAAGAGCTGGTG	660
Qy	661	atttgaagcaggggaagaactggcgatttggaaagcaggggaagaagaactggagatcgag	720
Db	661	ATTTTGAAGCAGGGGAAGAACTGGCGATTTGAAGCAGGGGAAGAACTGGAGATGCGG	720
Qy	721	aaactgaagaagaggaactggagatgcggaaactggaaatggagcaactgtgtatgtag	780
Db	721	AAACTGAAGAAGGAGCAACTGGAGATGCGGAACCTGAAATGGAGCAACTGTGTATGTAG	780
Qy	781	acacagaagatagttcagctgatgagcagaaaaagtcacatgttctgctcaagaataatg	840
Db	781	ACACAGAAGATAGTTTCAGCTGATGAGCAGAAAAAAGTACATGTTCTGCTCAAGAAAAATG	840
Qy	841	tacaacttcggatagtaatgatgcctcttggaaagtatttggataaaagataaattt	900
Db	841	TACAACCTGCCATAGTAAATGATGCCCTTTTGGAAAGTATTTTGGATTAAGATATAATTT	900
Qy	901	ttgatcatattaaagatttcgagccactattcgaacaaattggtgcgggtactgctaacc	960
Db	901	TTGATCATATTAAAGATTTTCGAGCCACTATTTCGAACAAATTTGTGCGGGTACTGCTAAAC	960
Qy	961	atggttacgggacaagaattgccaatgaacacctgtaccattaccagtgcagaagagcccg	1020
Db	961	ATGTTACGGGACAAGAAATTTGCCAATGAACCTGTACCAATTTACCATTTGCGCAGAGACCGC	1020
Qy	1021	cgaagctaccagcgaagaatttagatgccactccagagatgacttcgattagatgta	1080
Db	1021	CGAAGCTACCAGCGAAGAATTTAGATGCCACTCCAGAGATGACTTCGATTAGATGTTA	1080

Db	1021	CCCAAGTACCAGCGGAGAAATTAGATGCCACTCCAGAGGATGACTTCGCATTAGATGTTA	1080
Qy	1081	cagaatctcccagagaagtaqaattagattagatgaagagcaactgaagaagaatcaa	1140
Db	1081	CAGAACTCCCGAGGAAGTAGAATTAGTATTAGTGAAGGCAACTGAAGAAGTAACAA	1140
Qy	1141	cggaaagtggaccacggaagaagacccacccgaagaatttagatgccactccagaggatg	1200
Db	1141	CGGAAGTGGACCAACGGAAGACCAACCGAAGAAATTAGTGCCACTCCAGAGGATG	1200
Qy	1201	gatttcgcattagacgaactgcagaaggaagaacacagaagaacgtagagggagaagaaa	1260
Db	1201	GATTTCGCATTAGACGAAACTGCAGAAGGAGAAACAGAAAGTAGAGGGAGAGAAA	1260
Qy	1261	cagaagaagctgcagaaggaagaattatcagaagaactccagaaggaagaagaagtttag	1320
Db	1261	CAGAAGAAGCTGCAGAACGAGAAGTATCAGAAGAAACTTCCAGAGGAGAAAGAGTTAG	1320
Qy	1321	aggcaactcccagagatgatttcgcatttagatggaactacattagaagaacccgaagaaa	1380
Db	1321	AGGCAACTCCAGAGGATGATTTCGCATTAGTGGAACTACATTAGAAGAAACCCGAAGAA	1380
Qy	1381	ctgcagaaggaagaacccgttagagggagaagaacccgttagagggagaagaacccgttag	1440
Db	1381	CTGCAGAAGGAGAAACCCGTAGAGGGAGAGAAACCCGTAGAGGGAGAAACCCGTAG	1440
Qy	1441	agggaagaagctgcagaaggaagaagaggttagagggcaactccagagatgacttcc	1500
Db	1441	AGGGAGAAGAAGCTGCACAGAGAGAGAAGAGTTAGAGGCAACTCCACAGAGATGACTTCC	1500
Qy	1501	aattagaagaaccatcaggaaggaagaagggaagggaaggaaggaaggaaggaag	1560
Db	1501	AATTAGAAGAACCATCAGGAGAAGGAGGGAAGGGAAGAGAGGAGGGAAGGAGAGAG	1560
Qy	1561	gagaagccttagtcagcagtgccagtagtgcgcgaacccgttagaagtagtactcctctc	1620
Db	1561	GAGAAGCCTTTAGTAGCAGTGCCAGTAGTGGCGAAACCCGTAGAGTAGTACTCTCTGCTC	1620
Qy	1621	agcctgtcaaacccaaatgctcctcaacggcagatgaaactttattcgttatcttag	1680
Db	1621	AGCCTGTCAAAACCAATGCTCGCTCCAACGGCAGATGAACCTTATTCGTTGATATCTTAG	1680
Qy	1681	ataacgatttaacgtatgcagacattacatcctttgagccattattaaacaactccca	1740
Db	1681	ATAACGATTTAACGTTATGCAGACATFATACCTCTTTGAGCCATTATTTAAACAATCTCTCA	1740
Qy	1741	aggatccctgatlqagagagagcgtgtaacagttaccattcaaaaggaagcacctgtacaagtac	1800
Db	1741	AGGATCCTGTACGAGAGAGGCTGTAAACAGTACCATTCAAAAGAACACCTGTACAAGTAC	1800
Qy	1801	cagtggcagtagggccgcgcaagaagtgcacaacggagaagtgtgatccaactccaagagg	1860
Db	1801	CAGTGGCAGTAGGGCCCGCCGAAGAAGTCCCAACGGAAGAAATTGATGCAACTTCCAAGAGG	1860
Qy	1861	acgatttcgaattagaagaactgcagaagctccagagaagaagaatttagtattagaag	1920
Db	1861	ACGATTTCGAATTAGAAGAACTGCAGAAAGCTCCAGAGAGGAGGAAGAAATTAGTATTAGAAG	1920
Qy	1921	gagaagggagaacccaacggaagaagccaagaagagagagccaacagaagaagtggaagtg	1980
Db	1921	GAGAAGGAGAACCAACGGAAGAAGAGCCCAAGAGAAGGAGAGCAACAGAAGGAGAAGTGC	1980
Qy	1981	cagaagaagaatttagaggcaactccagaggacgttccgaattagaagaacccaacagag	2040
Db	1981	CAGAAGAAGAATTAGAGGCAACTCCAGAGGACGATTTCCGAATTAGAAGAAACCAACAGGAG	2040
Qy	2041	aagaagtagaagaacccgttagagggcgaagaactgcagaaggaagaagaagtggaagagg	2100
Db	2041	AAGAAGTAGAAGAAACCCGTAGAGGGCGAAGAAACTGCAGAGAGGAAGAAGTGGAGAGG	2100
Qy	2101	tacctgcagaagtagaagaagtgggaagagggtacctgcagaagtagaagaagtgggaagagg	2160
Db	2101	TACCTGCAGAACTAGAACAAGTGAAGAGGTACCTGCAGAACTAGAGAGTGGAGAGG	2160

Db 541 AAGAGCTGGCGATATGGAAGCAGGGGAGAAAGCTGGTGAATTTGGAACAGCGGGGAAGAAA 600
Qy 601 ctggcgatttgaagcaggggaagaaactggcgatttggaaagcaggggaagaaactggtg 660
Db 601 CTGGCGATTTGAAGCAGGGGAGAAACTGGCGATTTGAAGCAGGGGAGAAAGCTGGTG 660
Qy 661 atttgaagcaggggaagaaactggcgatttggaaagcaggggaagaaactggagatcg 720
Db 661 ATTtTGAAGCAGGGGAGAAACTGGCGATTTGGAAAGCAGGGGAGAAACTGGAGATGGG 720
Qy 721 aaactgaagaagagcaactgagatgcgaaactgaaatggaactgtgtatgtag 780
Db 721 AAACtGAAGAAGAGCAACTGGAGATGCGAAACTGAAATGGAGCAACTGTGTATGTAG 780
Qy 781 acacagaagatagttcagctgagcagaaagtagatgttctctctcaagaaaatg 840
Db 781 ACACAGAAGATAGTTCAGCTGATGAGCAGAAAAAAGTACATGTTCTCTCAAGAAAAATG 840
Qy 841 tacaacctgcgatagtaataatgacctcttggaaagtatttggataaaagataaatt 900
Db 841 TACAACCTGCCGATAGTAATGATGCCCTCTTTGGAAGTATTTTGGATAAAGATATAATTT 900
Qy 901 ttgatcatataaagatctcagccactattcgaaacaaatgtggcggtactgctaacc 960
Db 901 TTGATCATATTAAGATTTTCAGCCCACTATTTCGAACAATTTGGCGGGTACTGCTAACC 960
Qy 961 atgttacgggacaagaattggcaatgaaacctgtaccattaccagtggaagagcccg 1020
Db 961 ATGTTACGGGACAAGAATTTGCAATGAAACCTGTACCATTACCAGTGGCAGAAAGCCCG 1020
Qy 1021 ccaagtagtaccagcgaagaattagatgccactccagagagtagtactcgcattagatgta 1080
Db 1021 CCAAGTAGTACCAGCGGAAGAATTTAGATGCCACTCCAGAGAGTAGTTCGCATTAGATGTTA 1080
Qy 1081 cagaactctccgaggaagttagaatttagtattagatgaagagcaactgaagaagaatcaa 1140
Db 1081 CAGATCTCCCGAGGAGTAGAATTTAGTATTAGATGAGAGCAACTGAAGAGAAATCAA 1140
Qy 1141 cggaaagtgggaccacggaaggaagaccacccgaagaattagatgccactccagagatg 1200
Db 1141 CGGAAGTGGGACCAACGGGAAGAGACCAACCGAAGAATTAGATGCCACTCCAGAGGATG 1200
Qy 1201 gatttcgcattagcgaactcgaagagagaaacagagaacatagaggagaagaaa 1260
Db 1201 GATTTCGCATTAGACGAACCTCGAAGAGAGAAACAGAGAACCTAGAGGGAGAGAAA 1260
Qy 1261 cagaagaactcgaagaggaagtatcagaagaactccgaagaggaagaagagattag 1320
Db 1261 CAGAAGAAGCTCCAGAGGAGAGAGTATCAGAGAAACTCCAGAGGAGAGAAGAGTAG 1320
Qy 1321 aggcaactccagaggatgatttcgcattagatggaactacattagaagaaccggaagaa 1380
Db 1321 AGGCCAACTCCAGAGGATGATTTCCGATTAGATGGAACCTACATTAGAAGAAACCCGAAGA 1380
Qy 1381 ctgcagaaggaagaaccgtagagggaagaacccgtagaggaggaagaaccgtag 1440
Db 1381 CTGCAGAAGGAGAGAAACCCGTAGAGGGAGAGAAACCCGTAGAGGGAGAGAAACCCGTAG 1440
Qy 1441 agggagaagaactcgaagagagaagaagtagtagagcaactccagagagatgacttcc 1500
Db 1441 AGGGAGAGAAGACTCCAGAGAGAGAGAGAGATTAGAGCAACTCCAGAGGATGACTTCC 1500
Qy 1501 aattagaagaacctcaggagaaggaaggggaaggggaaggggaaggggaaggaag 1560
Db 1501 AATTAGAAGAACCATCAGGAGAAGGAGAAGGGGAAGGAGCAAGCAAGGGGAAGGAGAAG 1560
Qy 1561 gagaagcgttagtagcagtagtgcagtagtgcgaacccggtagaagtagtactctgtc 1620
Db 1561 GAGAAGCCGTAGTAGCAGTGGCCAGTAGTGGCCGAACCCGGTAGAAGTAGTACTCTGTCTC 1620
Qy 1621 agcctgtcaaaccaatggctgcctcaacggcagatgaaacttattcgttgatatcttag 1680
Db 1621 AGCCTGTCAAACCAATGGTCTGCTCCACGGCAGATGAACACTTATTCTGTTGATATCTTAG 1680

Qy 1681 ataacgatttaacgtatgcagacattacatcctttagccattatttataacaaactcctca 1740
Db 1681 ATAACGATTTAACGTATGCAGACATTTACATCCTTTTGAGCCATTATTTAAACAAATCCTCA 1740
Qy 1741 aggatccttgatgcagagagagcgttaacagtagcattacataaaggaagcaactgtacaagtac 1800
Db 1741 AGGATCCTTGATGCAGGAGAGGCTGTAAACAGTACCATCAAGGAAGACCTGTACAAAGTAC 1800
Qy 1801 cagtggcagtagggcccgcaagaagtgcacaacgaagaaattgatgcaactcccaagagg 1860
Db 1801 CAGTGGCAGTAGGGCCCGCAAGAAAGTGGCCACCGAAGAAATTGATGCAACTCCCAAGAGG 1860
Qy 1861 acgatttcgaattagaagaagcagcagagctccagaggaaggaataattagattagaag 1920
Db 1861 ACGATTTCCGAATTTAGAAGAACTGCAGAAGCTCCAGAGAGAGGAGAATTAGTATTAGAAG 1920
Qy 1921 gagaagagaaccaacggaagaagccaagaagaagagccaagaagagagccaagaagagagtc 1980
Db 1921 GAGAAGGAGAACCACGGAAGAAAGCAAGAAAGAGAGCCAAACAGAAAGGAGAAAGTGC 1980
Qy 1981 cagaagaagaattagagcgaactccagagagcagatttcgaaattagaagaaccaaagag 2040
Db 1981 CAGAAGAAGAATTAGAGGCAACTCCAGAGGAGGATTTCSAATTTAGAAGAACCAACAGGAG 2040
Qy 2041 aagaagttagaagaaccctagaggcggaagaaaactcggaaggaaggaagaagtggaaaggg 2100
Db 2041 AAGAAGTAGAAGAAACCTAGAGGGCGAAGAAACTGCAGAGAGGAGAAGAACTGGAAAGG 2100
Qy 2101 tactctcagaagttagaagaagtgcgaagaggtacctcgagaagtagaagaagtgcgaagagg 2160
Db 2101 TACCTGCAAGAGTAGAAGAGTGGAAAGGTACCTTCGCAAGTAGAAGAAAGTGGAAAGAGG 2160
Qy 2161 taccagaagaagtagaagaggtaccgcgagaagtagaagaagtggaagaggtaccagaag 2220
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCCGAGAGTAGAAGAACTGGAAGAGGTACCAGAAG 2220
Qy 2221 aagtggaaagaggtaccagaagaagtggaaaggtaccggaaggaagtggaaagaggtaccag 2280
Db 2221 AAGTGGAAAGAGGTACCAGAAGAAGTGGAAAGGTACCAGAAAGTGGAAAGAGGTACCAG 2280
Qy 2281 aagaagtggaaagaagtgcgaagaagttagaagaagttagaggtaccgcggttagtagaagttag 2340
Db 2281 AAGAAGTGGAAAGAGTGGAGAGAGTAGAAGAAAGTAGAGGTACCAGCGGTAGTAGAAGTAG 2340
Qy 2341 aagtaccgcggttagtagaagaagaggtgcgaagaagtagaagaagaagaagaagagagg 2400
Db 2341 AAGTACCAGCGGTAGTAGAAGAAGAGGTGCCAGAAGAAAGTAGAAGAAGAAAGAAAGAGG 2400
Qy 2401 aagaaccagtagagaagaagaagtattacaaattagtaataaccatcggaagaagaatatac 2460
Db 2401 AAGAACCACTAGAGGAAGAAGATGTTACAAATTAGTATTACCATCGGAAGAAGATATAC 2460
Qy 2461 aattagacaacaaagaaacgaataggtctctggaaatttactatcatcgacatgc 2520
Db 2461 AATTAGACAACCAAGAAAGACGAATTAGGCTCTGGAAATTTTATCTATCATCGACATGC 2520
Qy 2521 actaccaagaagttcccaaggaatttaggaagaagaagaactcagtgatccat 2580
Db 2521 ACTACCAAGACGTTCCAAAGGAATTTATGGAAGAAAGAAAGAAACTGCAGTGTATCCAT 2580
Qy 2581 tgaaccagaagatttgcgaaggaagattcaaatctacagaatggctcaacatcttc 2640
Db 2581 TGAACCAAGAGATTTTCAAGAGGAAGATTTCAACATCTACAGAATGGCTACATTCATTC 2640
Qy 2641 aagccttagaagggcagctggaaacatttagaagtgccttaataagcttagaagaagat 2700
Db 2641 AAGGCCTAGAAGGCCAGCTGGGAACGATTAGAAGTAGAGCTTAATTAAGGCTAGAGAAAGAT 2700
Qy 2701 ggaatgaagaagaataaagaatgggctgcgtcttcgcgttaattgaaaaataaattggt 2760
Db 2701 GGATGGAACAAAGAAATAAGAAATGGGCTGGCTTGCCTTAAATTGAAAAATAAATGCT 2760

```
Oy 2761 cagaataagtcagaaatcacaagaaagagagccagctggttgagaaacagagagt 2820
Db 2761 CAGATATAGTCAATTTCAACAAAGGAAGAGCCAGCTGGTTTGAGAAACAGAGAT 2820
Oy 2821 ggaagcagagaaatggaagaaatggtttaagcagagtcacaaatcccaaatgattcac 2880
Db 2821 GGAGCGAGGAGAAATGGAAAAATGGTTTAAAGCAGAGTCAAAATCCCAAAATGATTCAC 2880
Oy 2881 actgaaaaatgatgaacgacactcactcaattatttatttataaattcttgtgaagata 2940
Db 2881 ACTTGAATAAATGGATGAACGACACTCAATCCAAATTTATTAAATTTCTGTGAAGATA 2940
Oy 2941 tgtcacaaattgaaaaaagaaacacaaagaaatggttaataatcactggaagaaagacg 3000
Db 2941 TGTCACAATTTGAAACAAAGAAACCAAGAAATGTTTAATCAATCACTGGAAAAAGAACG 3000
Oy 3001 aacggggttatgggttctgaacatttggaagttagaccacatcaaaattatttaattgtg 3060
Db 3001 AACGGGGTTATGGTCTGAATCATTTTGAAGTTATGACCACATCAAAATTTATTAATGTGG 3060
Oy 3061 ctaagagtcgagaatggtaccgtgccaatcctaataataatagagaagagagaactca 3120
Db 3061 CTAGAGTCGAGATGTTACCGTGCCCAATCTATATATAATAGAGAAAGAGAACTCA 3120
Oy 3121 tgaatggttctcctcaaaaagaaacgaatatttaggacaaagaaatggaagaaatggact 3180
Db 3121 TGAATGTTTTCTCTAAACAAAGCAATATTTAGGACAAAGAAATGGAATAATGGACT 3180
Oy 3181 catggaaaaagttataatttttgggttcacaaatggtacacaaatcttctggaaaaa 3240
Db 3181 CATTCGAAAAAAGTTAAATTTTGGTTCATTTCAATGTTGTACAAACATTTTCTGGAAA 3240
Oy 3241 cgcctaaccaaggaagaaatgaatcaatttggtaataataataaagttgtaattatagaa 3300
Db 3241 CGCCTAACCAAGGAAGATGAATCAATTTGTTAATGAATAAAGATTTGAATTATAGAA 3300
Oy 3301 aaagaaacagattattctcttataaaaaataaattc 3337
Db 3301 AAAAGAACAGATTATCTCTTATAAAATAAATAATTC 3337
```

```
RESULT 3
AC067910/c
LOCUS
DEFINITION Homo sapiens chromosome 22 clone RP11-51H9 map 22, LOW-PASS
SEQUENCE SAMPLING.
AC067910
VERSION AC067910.1 GI:7652049
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 70206)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgaiter,B., Brown,A., Burkett,G.,
Campolano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collins,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hegos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lileu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPeeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pterre,N.,
```

TITLE
JOURNAL
COMMENT

Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L7007
Center clone name: 51_R_9

* NOTE: This record contains 87 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

1 736: contig of 736 bp in length
* 737 836: gap of 100 bp
* 837 1561: contig of 725 bp in length
* 1562 1661: gap of 100 bp
* 1662 2378: contig of 717 bp in length
* 2379 2478: gap of 100 bp
* 2479 3202: contig of 724 bp in length
* 3203 3302: gap of 100 bp
* 3303 4006: contig of 704 bp in length
* 4007 4106: gap of 100 bp
* 4107 4798: contig of 692 bp in length
* 4799 4898: gap of 100 bp
* 4899 5611: contig of 713 bp in length
* 5612 5711: gap of 100 bp
* 5712 6416: contig of 705 bp in length
* 6417 6516: gap of 100 bp
* 6517 7213: contig of 697 bp in length
* 7214 7313: gap of 100 bp
* 7314 8017: contig of 704 bp in length
* 8018 8117: gap of 100 bp
* 8118 8833: contig of 716 bp in length
* 8834 8933: gap of 100 bp
* 8934 9656: contig of 723 bp in length
* 9657 9756: gap of 100 bp
* 9757 10472: contig of 716 bp in length
* 10473 10572: gap of 100 bp
* 10573 11284: contig of 712 bp in length
* 11285 11384: gap of 100 bp
* 11385 12108: contig of 724 bp in length
* 12109 12208: gap of 100 bp
* 12209 12910: contig of 702 bp in length
* 12911 13010: gap of 100 bp
* 13011 13710: contig of 700 bp in length
* 13711 13810: gap of 100 bp
* 13811 14500: contig of 690 bp in length
* 14501 14600: gap of 100 bp
* 14601 15305: contig of 705 bp in length
* 15306 15405: gap of 100 bp
* 15406 16114: contig of 709 bp in length
* 16115 16214: gap of 100 bp
* 16215 16914: contig of 700 bp in length

* 16915 17014: gap of 100 bp
* 17015 17722: contig of 708 bp in length
* 17723 17822: gap of 100 bp
* 17823 18517: contig of 695 bp in length
* 18518 18617: gap of 100 bp
* 18618 19329: contig of 712 bp in length
* 19330 19429: gap of 100 bp
* 19430 20132: contig of 703 bp in length
* 20133 20232: gap of 100 bp
* 20233 20955: contig of 723 bp in length
* 20956 21055: gap of 100 bp
* 21056 21762: contig of 707 bp in length
* 21763 21862: gap of 100 bp
* 21863 22560: contig of 698 bp in length
* 22561 22660: gap of 100 bp
* 22661 23351: contig of 691 bp in length
* 23352 23451: gap of 100 bp
* 23452 24145: contig of 694 bp in length
* 24146 24245: gap of 100 bp
* 24246 24952: contig of 707 bp in length
* 24953 25052: gap of 100 bp
* 25053 25756: contig of 704 bp in length
* 25757 25856: gap of 100 bp
* 25857 26568: contig of 712 bp in length
* 26569 26668: gap of 100 bp
* 26669 27391: contig of 723 bp in length
* 27392 27491: gap of 100 bp
* 27492 28187: contig of 696 bp in length
* 28188 28287: gap of 100 bp
* 28288 28993: contig of 706 bp in length
* 28994 29093: gap of 100 bp
* 29094 29799: contig of 706 bp in length
* 29800 29899: gap of 100 bp
* 29900 30629: contig of 730 bp in length
* 30630 30729: gap of 100 bp
* 30730 31418: contig of 689 bp in length
* 31419 31518: gap of 100 bp
* 31519 32200: contig of 682 bp in length
* 32201 32300: gap of 100 bp
* 32301 33002: contig of 702 bp in length
* 33003 33102: gap of 100 bp
* 33103 33782: contig of 680 bp in length
* 33783 33882: gap of 100 bp
* 33883 34581: contig of 699 bp in length
* 34582 34681: gap of 100 bp
* 34682 35382: contig of 701 bp in length
* 35383 35482: gap of 100 bp
* 35483 36178: contig of 696 bp in length
* 36179 36278: gap of 100 bp
* 36279 36998: contig of 720 bp in length
* 36999 37098: gap of 100 bp
* 37099 37805: contig of 707 bp in length
* 37806 37905: gap of 100 bp
* 37906 38620: contig of 715 bp in length
* 38621 38720: gap of 100 bp
* 38721 39424: contig of 704 bp in length
* 39425 39524: gap of 100 bp
* 39525 40243: contig of 719 bp in length
* 40244 40343: gap of 100 bp
* 40344 41048: contig of 705 bp in length
* 41049 41148: gap of 100 bp
* 41149 41851: contig of 703 bp in length
* 41852 41951: gap of 100 bp
* 41952 42658: contig of 707 bp in length
* 42659 42758: gap of 100 bp
* 42759 43478: contig of 720 bp in length
* 43479 43578: gap of 100 bp
* 43579 44282: contig of 704 bp in length
* 44283 44382: gap of 100 bp
* 44383 45080: contig of 698 bp in length
* 45081 45180: gap of 100 bp
* 45181 45894: contig of 714 bp in length
* 45895 45994: gap of 100 bp

* 45995 46707: contig of 713 bp in length
* 46708 46807: gap of 100 bp
* 46808 47546: contig of 739 bp in length
* 47547 47646: gap of 100 bp
* 47647 48350: contig of 704 bp in length
* 48351 48450: gap of 100 bp
* 48451 49166: contig of 716 bp in length
* 49167 49266: gap of 100 bp
* 49267 49978: contig of 712 bp in length
* 49979 50078: gap of 100 bp
* 50079 50785: contig of 707 bp in length
* 50786 50885: gap of 100 bp
* 50886 51582: contig of 697 bp in length
* 51583 51682: gap of 100 bp
* 51683 52384: contig of 702 bp in length
* 52385 52484: gap of 100 bp
* 52485 53196: contig of 712 bp in length
* 53197 53296: gap of 100 bp
* 53297 54015: contig of 719 bp in length
* 54016 54115: gap of 100 bp
* 54116 54833: contig of 718 bp in length
* 54834 54933: gap of 100 bp
* 54934 55620: contig of 687 bp in length
* 55621 55720: gap of 100 bp

Query Match 1.2%; Score 40; DB 2; Length 70206;
Best Local Similarity 100.0%; Pred. No. 3.3e-08;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1527 gaaggggaagagagagagagagagagagagagagagag 1566
|||||
Db 24332 GAAGGGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 24293

RESULT 4

AC027625 AC027625 165868 bp DNA linear HTG 09-AUG-2001
LOCUS Homo sapiens chromosome 4 clone RP11-27013, WORKING DRAFT SEQUENCE,
DEFINITION 2 unordered pieces.

ACCESSION AC027625.4 GI:14333966
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens

REFERENCE 1 (bases 1 to 165868)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 165868)
AUTHORS Waterston,R.H.
TITLE Direct Submission

JOURNAL Submitted (30-MAR-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT On Jun 9, 2001 this sequence version replaced gi.9838122.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H_NH0270103
----- Summary Statistics -----
Sequencing vector: M13; 40%
Chemistry: Dye-primer ET; 40% of reads
Chemistry: Dye-terminator Big Dye; 60% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 165641 bases at least Q40
Consensus quality: 165729 bases at least Q30

Consensus quality: 165743 bases at least Q20
 Insert size: 168000; agarose-fp
 Insert size: 166538; sum-of-contigs
 Quality coverage: 11.53 in Q20 bases; agarose-fp
 Quality coverage: 11.63 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 57897: contig of 57897 bp in length
 * -57898 57997: gap of unknown length
 * 57998 165868: contig of 107871 bp in length.

FEATURES

source
 Location/Qualifiers
 1..165868
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="4"
 /clone="RP11-27013"

misc_feature

1..57897
 /note="assembly_name:Contig5
 clone_end:SP6
 vector_side:right"

misc_feature

57998..165868
 /note="assembly_name:Contig6
 clone_end:T7
 vector_side:left"

BASE COUNT 48495 a 32516 c 32464 g 52293 t 100 others
 ORIGIN

Query Match 1.2%; Score 40; DB 2; Length 165868;
 Best Local Similarity 100.0%; Pred. No. 2.9e-08;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1527 gaaggggaaggaaggaaggaaggaaggaaggaaggaag 1566
 |||
 Db 3740 GAAGGGGAAGGAGAGGAGAGGAGGAGAGGAGAGGAGAG 3701

RESULT 5
 AL451079 186562 bp DNA linear HTG 05-OCT-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-307017, *** SEQUENCING IN
 DEFINITION PROGRESS ***, 2 unordered pieces.
 ACCESSION AL451079 AC025875
 VERSION AL451079.6 GI:15983906
 KEYWORDS HTG; HTGS_PHASE1; HTGS_ACTIVEPIN; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE human.

ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 186562)
 AUTHORS Williams,S.
 TITLE Direct Submission
 JOURNAL Submitted (04-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Oct 8, 2001 this sequence version replaced gi:15591231.
 Draft Sequence Produced by Whitehead Institute/MIT Center for
 Genome Research, 320 Charles Street,
 Cambridge, MA 02141, USA
 http://www-seq.wi.mit.edu
 ----- Genome Center
 Center: Sanger Centre
 Web site: http://www.sanger.ac.uk
 Contact: humquery@sanger.ac.uk
 ----- Project Information

COMMENT

Center project name: bA307017
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: M13; M7815; 26% of reads
 Chemistry: Dye-terminator Big Dye; 108752; 73% of reads
 Consensus quality: 185620 bases at least Q40
 Consensus quality: 185985 bases at least Q30
 Consensus quality: 186189 bases at least Q20
 Insert size: 18462; sum-of-contigs
 Insert size: 172836; 9.6% error; agarose-fp
 Quality coverage: 12.37x in Q20 bases; sum-of-contigs Quality
 coverage: 13.42x in Q20 bases; agarose-fp

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 2201: contig of 2201 bp in length
 * 2202 2301: gap of 100 bp
 * 2302 186562: contig of 184261 bp in length.

FEATURES

source
 Location/Qualifiers
 1..186562
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-307017"
 /clone_lib="RPCI-11.2"

misc_feature

1..2201
 /note="assembly_fragment:00529"
 2302..186562
 /note="assembly_fragment:05314
 clone_end:T7
 vector_side:right"

BASE COUNT 50258 a 43749 c 43919 g 48534 t 102 others
 ORIGIN

Query Match 1.0%; Score 34; DB 2; Length 186562;
 Best Local Similarity 100.0%; Pred. No. 2e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1533 gaaggaaggaaggaaggaaggaaggaaggaag 1566
 |||
 Db 27017 GAAGGAGAGGAGAGGAGGAGGAGGAGGAGGAGGAGAG 27050

RESULT

6
 AC096541/C 200898 bp DNA linear HTG 18-SEP-2001
 LOCUS Homo sapiens chromosome 1 clone RP11-330M19, WORKING DRAFT
 DEFINITION SEQUENCE, 4 unordered pieces.
 ACCESSION AC096541 AL359084
 VERSION AC096541.1 GI:15638689
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 200898)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 200898)
 AUTHORS Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J. and Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (18-SEP-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA

COMMENT

On Sep 18, 2001 this sequence version replaced gi:12331065.

----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgctgs@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP11-330M19 (sc0667)
 ----- Summary Statistics

Sequencing vector: plasmid; 38% of reads
 Sequencing vector: plasmid; 108752; 62% of reads
 Chemistry: Dye-terminator ET; 89% of reads
 Chemistry: Dye-terminator Big Dye; 11% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 199175 bases at least Q40
 Consensus quality: 199939 bases at least Q30
 Consensus quality: 200341 bases at least Q20
 Insert size: 200598; sum-of-contigs
 Quality coverage: 9.5x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 2137: contig of 2137 bp in length
 * 2138 2237: gap of unknown length
 * 2238 29137: contig of 26900 bp in length
 * 29138 29237: gap of unknown length
 * 29238 55080: contig of 25843 bp in length
 * 55081 55180: gap of unknown length
 * 55181 200898: contig of 145718 bp in length.

FEATURES

source
 1. .200898
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /clone="RP11-330M19"
 /clone_lib="RPCI human BAC library 11"
 misc_feature
 1. .2137
 /note="assembly_name:Contig18"
 misc_feature
 2238. .29137
 /note="assembly_name:Contig19"
 misc_feature
 29238. .55080
 /note="assembly_name:Contig20"
 misc_feature
 55181. .200898
 /note="assembly_name:Contig21"
 BASE COUNT 50081 a 49279 c 48900 g 52334 t 304 others
 ORIGIN

Query Match 1.0%; Score 34; DB 2; Length 200898;
 Best Local Similarity 100.0%; Pred. No. 2e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1533 gaaggagaggaaggggaaggaaggaaggaag 1566
 |||||
 Db 55487 GAAGGAGAGGAGAGGGGAGGAGGAGGAGAG 55454

RESULT 7
 AL645546/c AL645546 225609 bp DNA linear HTG 13-NOV-2001
 LOCUS Mus musculus chromosome 13 clone RP23-124G18, *** SEQUENCING IN
 DEFINITION PROGRESS ***, in unordered pieces.
 ACCESSION AL645546
 VERSION AL645546.4 GI:16944328
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE house mouse.

ORGANISM

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (sites)
 Sims, S.
 Direct Submission
 Submitted (08-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
 On Nov 15, 2001 this sequence version replaced gi:16754763.

COMMENT

----- Genome Center
 Center: UK Medical Research Council
 Center code: UK-MRC
 Web site: <http://mrcseq.har.mrc.ac.uk>
 Contact: mouseq@har.mrc.ac.uk
 ----- Project Information
 Center project name: bml24G18
 ----- Summary Statistics
 Assembly program: XGAP4; version 4.5
 Sequencing vector: plasmid; 108752; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Consensus quality: 224286 bases at least Q40
 Consensus quality: 224563 bases at least Q30
 Consensus quality: 224933 bases at least Q20
 Insert size: 225409; sum-of-contigs
 Insert size: 216124; 7.8% error; agarose-fp
 Quality coverage: 11.91x in Q20 bases; sum-of-contigs Quality
 coverage: 13.08x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES

source
 1. .225609
 Location/Qualifiers
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="13"
 /clone="RP23-124G18"
 /clone_lib="RPCI-23"
 1. .101657
 /note="assembly_fragment:01246"
 clone_end:T7
 vector_side:left
 misc_feature
 101758. .190829
 /note="assembly_fragment:02144.0"
 misc_feature
 190930. .225609
 /note="assembly_fragment:03429"
 clone_end:SP6
 vector_side:right
 BASE COUNT 66198 a 45409 c 44739 g 69047 t 216 others
 ORIGIN

Query Match 1.0%; Score 34; DB 2; Length 225609;
 Best Local Similarity 100.0%; Pred. No. 1.9e-05;
 Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1533 gaaggagaggaaggggaaggaaggaaggaag 1566
 |||||
 Db 101565 GAAGGAGAGGAGAGGGGAGGAGGAGGAGAG 101532

RESULT

AX251053
 LOCUS AX251053 6862 bp DNA linear PAT 05-OCT-2001
 DEFINITION Sequence 21 from Patent WO0168912.
 ACCESSION AX251053
 VERSION AX251053.1 GI:15984476
 KEYWORDS synthetic construct.
 SOURCE synthetic construct
 ORGANISM artificial sequence.

repeat_region	2667. .2733	/note="MER21B d"
repeat_region	2737. .3281	/note="(AT) d"
repeat_region	3303. .3658	/note="L1PA7 c"
repeat_region	3663. .3760	/note="L1PA2 d"
repeat_region	4290. .4490	/note="HERVL d"
repeat_region	4701. .5003	/note="X20"
repeat_region	5120. .5260	/note="L1PA2 d"
repeat_region	5267. .5702	/note="L1P5A1 c"
repeat_region	5799. .5939	/note="L1P_MA2 c"
repeat_region	5990. .6123	/note="L1MD2 c"
repeat_region	6515. .6953	/note="AluJo d"
repeat_region	6962. .7075	/note="L1P5A1 c"
repeat_region	7123. .7183	/note="L1MD2 c"
repeat_region	7200. .7402	/note="L1MA10 c"
repeat_region	7403. .7868	/note="L1MB7 c"
repeat_region	7869. .8160	/note="L1MA9 c"
repeat_region	8165. .8248	/note="MER28 c"
repeat_region	8249. .9160	/note="TIGGER2 c"
repeat_region	9161. .9833	/note="L1P_MA2 c"
repeat_region	9857. .9900	/note="L1P30 c"
repeat_region	9939. .11947	/note="L1P_MA2 c"
repeat_region	11949. .12598	/note="L1P_MA2 c"
repeat_region	12606. .13027	/note="L1P_MA2 c"
repeat_region	13032. .13675	/note="L1P_MA2 c"
repeat_region	13701. .14702	/note="L1P_MA2 c"
repeat_region	14859. .15042	/note="L1P_MA2 c"
repeat_region	15082. .15964	/note="L1P_MA2 c"
repeat_region	15965. .16512	/note="L1PA7 c"
repeat_region	16513. .18088	/note="L1 c"
repeat_region	18637. .18842	/note="L1 d"
repeat_region	18851. .19123	/note="L1MD2 c"
repeat_region	19127. .19334	/note="AluJb c"
repeat_region	19472. .19602	/note="L1MD2 c"
repeat_region	19610. .19995	/note="L1MA2 c"
repeat_region	19996. .20602	/note="MSTA c"
repeat_region	20607. .20882	/note="L1MA2 c"
repeat_region	21614. .21910	/note="L1MB7 c"
repeat_region	21925. .22090	/note="L1MA2 c"
repeat_region	22397. .22528	/note="X25"
repeat_region	23028. .23108	/note="L1P_MA2 c"
repeat_region	23109. .23479	/note="L1 c"
repeat_region	23481. .24297	/note="L1 c"
repeat_region	24299. .25822	/note="L1 c"
repeat_region	25829. .26690	/note="L1P5A1 c"
repeat_region	26915. .27203	/note="L1P_MA2 c"
repeat_region	27206. .27535	/note="AluY c"
repeat_region	27538. .28856	/note="L1 c"
repeat_region	28860. .29705	/note="L1P_MA2 c"
repeat_region	29716. .29923	/note="L1MD2 c"
repeat_region	29934. .30218	/note="L1P5A1 c"
repeat_region	30226. .30307	/note="L1PBB_5 c"
repeat_region	30442. .30465	/note="L1PA16_5 c"
repeat_region	31382. .31556	/note="(A) d"
repeat_region	31726. .31912	/note="MER5A d"
repeat_region	32492	/note="L1MA2 c"
repeat_region	32933. .33064	/note="32667. .X08"
repeat_region	33072. .33390	/note="L1ME3A d"
repeat_region	33770. .34007	/note="L1ME4 d"
repeat_region	34042. .34078	/note="LINE2 c"
repeat_region	34083. .34374	/note="L1MA5 c"
repeat_region	34467. .35368	/note="LINE2 c"
repeat_region	35491. .36126	/note="LINE2 c"
repeat_region	36127. .36352	/note="L1ME3A d"
repeat_region	37739. .37981	/note="L1ME2 d"
repeat_region	38272. .38364	/note="MIR c"
repeat_region	39305. .39363	/note="MIR d"
repeat_region	39436. .39991	/note="MER511 d"
repeat_region	39992. .40087	/note="L1 c"
repeat_region	40088. .40978	/note="L1 d"
repeat_region	42134. .42302	/note="L1PA2 d"
repeat_region	44358. .44681	/note="L3 c"
repeat_region	46558. .46702	/note="MSTA1 c"
repeat_region	46720. .46960	/note="LINE2 c"
repeat_region		/note="L1PA7 d"

Query Match 1.0%; Score 33; DB 9; Length 240000;
Best Local Similarity 100.0%; Pred. No. 5.7e-05;
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1527 gaagggaaggaaggaaggaaggaaggaagga 1559

Db 190008 GAAGGGAAGGAAGGAAGGAAGGAAGGAAGGA 190040

RESULT 12

AC096301

LOCUS

DEFINITION Rattus norvegicus chromosome Rf1 clone CH230-233D20, WORKING DRAFT HTG 20-DEC-2001

AC096301 SEQUENCE, 33 unordered pieces.

AC096301 AC096301.4 GI:17943998

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE Norway rat.

ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

1 (bases 1 to 246176)

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C., Alsbrooks,S.L., Amaralunga,H.C., Are,J.R., Banks,T., Barbara,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Dena,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homs,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,C., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H., Lozano,R., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S., Ogih,M., Okuwon,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,F., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shoohtari,N., Sisson,I., Sodergren,E., Sonaik,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Thomas,N., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williams,A., Wleczek,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D., Weinstein,G. and Gibbs,R.

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (17-SEP-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Dec 20, 2001 this sequence version replaced gi:17062231.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GEMU

Center clone name: CH230-233D20

----- Summary Statistics

Assembly program: Phrap; version 0.990329First call to

findPhrapList

Consensus quality: 227668 bases at least Q40

Consensus quality: 231229 bases at least Q30

Consensus quality: 233891 bases at least Q20

Estimated insert size: 229594; sum-of-contigs estimation

Quality coverage: 0x in Q20 bases; agarose-fp estimation

Quality coverage: 4.9x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

consists of 33 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as

runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

as soon as it is available and the accession number will

be preserved.

1 25258: contig of 25258 bp in length

25259 25358: gap of unknown length

25359 48246: contig of 22888 bp in length

48247 48346: gap of unknown length

48347 66685: contig of 18339 bp in length

66686 66785: gap of unknown length

66786 83731: contig of 16946 bp in length

83732 83831: gap of unknown length

83832 93990: contig of 15559 bp in length

93991 99491: gap of unknown length

110000: contig of 10510 bp in length

110101: gap of unknown length

122163: contig of 12063 bp in length

122164 122263: gap of unknown length

122264 134794: contig of 12531 bp in length

134795 134894: gap of unknown length

134895 146778: contig of 11884 bp in length

146779 146878: gap of unknown length

146879 155054: contig of 8176 bp in length

155055 155154: gap of unknown length

155155 162599: contig of 7445 bp in length

162600 172264: contig of 9565 bp in length

172265 172364: gap of unknown length

172365 180466: contig of 8102 bp in length

180467 180566: gap of unknown length

180567 186883: contig of 6317 bp in length

186884 186983: gap of unknown length

186984 193290: contig of 6307 bp in length

193291 193390: gap of unknown length

193391 200294: contig of 6904 bp in length

200295 200394: gap of unknown length

200395 205791: contig of 5397 bp in length

205792 205891: gap of unknown length

205892 210758: contig of 4767 bp in length

210759 215486: gap of unknown length

215487 215586: contig of 4728 bp in length

215587 220044: contig of 4458 bp in length

220045 220144: gap of unknown length

220145 223051: contig of 2907 bp in length

223052 223151: gap of unknown length

223152 225620: contig of 2469 bp in length

225621 225720: gap of unknown length

225721 228107: contig of 2387 bp in length

228108 228207: gap of unknown length

228208 228207: gap of unknown length

```
* 228208 231207: contig of 3000 bp in length
* 231208 231307: gap of unknown length
* 231308 232936: contig of 1629 bp in length
* 232937 233036: gap of unknown length
* 233037 234444: contig of 1408 bp in length
* 234445 234544: gap of unknown length
* 234545 236460: contig of 1916 bp in length
* 236461 236560: gap of unknown length
* 236561 238859: contig of 2299 bp in length
* 238860 238959: gap of unknown length
* 238960 240759: contig of 1800 bp in length
* 240760 240859: gap of unknown length
* 240860 242544: contig of 1685 bp in length
* 242545 242644: gap of unknown length
* 242645 243695: contig of 1051 bp in length
* 243696 243795: gap of unknown length
* 243796 244851: contig of 1056 bp in length
* 244852 244951: gap of unknown length
* 244952 246176: contig of 1225 bp in length.
FEATURES
    source
        Location/Qualifiers
            1..246176
                /organism="Rattus norvegicus"
                /db_xref="taxon:10116"
                /chromosome="p1"
                /clone="CH230-233D20"
BASE COUNT 65649 a 55496 c 55272 g 66413 t 3346 others
ORIGIN

Query Match      1.0%  Score 33:  DB 2:  Length 246176;
Best Local Similarity 100.0%:  Pred. No. 5.7e-05;
Matches 33:  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

QY 1534 aaggagaaggaggaagggaaggagaggaag 1566
|||||
Db 191928 AAGGAGAGGAGAGGAGGAGGAGGAGAG 191960

RESULT 13
AP001848
LOCUS
DEFINITION Homo sapiens chromosome 18 clone RP11-814018 map 18p11.2, WORKING
DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION AP001848
VERSION AP001848.3 GI:9757461
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens DNA, clone:RP11-814018.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Homo sapiens 224,010 genomic DNA of 18p11.2
Published Only in Database (2000) In press
2 (bases 1 to 224010)
Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
Direct Submission
Submitted (18-APR-2000) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail:hattori@gscc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel:81-42-778-9923, Fax:81-42-778-9924)
On Aug 9, 2000 this sequence version replaced gi:8117511.
----- Genome Center
Center: RIKEN Genomic Sciences Center(GSC)
Center code: RIKEN
Web site: http://hgp.gsc.riken.go.jp/
Contact: hattori@gscc.riken.go.jp
----- Project Information
Center project name: HumDraft18
Center clone name: RP11-814018

----- Summary Statistics
Sequencing vector: PCR products; 100% of reads
Chemistry: Dye-terminator ET-amersham; 100% of reads Assembly
Program: Phrap; version 0.990329
Consensus quality: 212358 bases at least Q40
Consensus quality: 215922 bases at least Q30
Consensus quality: 218536 bases at least Q20
Insert size: 220910; sum-of-contigs
Quality coverage: 9.03x in Q20 bases; sum-of-contigs
-----
NOTE: This is a 'working draft' sequence. It currently consists of
32 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.
1 15221 contig of 15221 bp in length
15322 30276 contig of 14955 bp in length
30377 40753 contig of 10377 bp in length
40854 53793 contig of 12940 bp in length
53894 65011 contig of 11118 bp in length
65112 75224 contig of 10113 bp in length
75325 84814 contig of 9490 bp in length
84915 96348 contig of 11434 bp in length
96449 106172 contig of 9724 bp in length
106273 114921 contig of 8649 bp in length
115022 123606 contig of 8585 bp in length
123707 131751 contig of 8045 bp in length
131852 138950 contig of 7099 bp in length
139051 144550 contig of 5500 bp in length
144651 152953 contig of 8303 bp in length
153054 159224 contig of 6171 bp in length
159325 165736 contig of 6412 bp in length
165837 173107 contig of 7271 bp in length
173208 178488 contig of 5281 bp in length
178589 183587 contig of 4999 bp in length
183688 187911 contig of 4224 bp in length
188012 192383 contig of 4372 bp in length
192484 196811 contig of 4328 bp in length
196912 201215 contig of 4304 bp in length
201316 205918 contig of 4603 bp in length
206019 210031 contig of 4013 bp in length
210132 213781 contig of 3650 bp in length
213882 217164 contig of 3283 bp in length
217265 219263 contig of 1999 bp in length
219364 220624 contig of 1261 bp in length
220725 221858 contig of 1134 bp in length
221959 224010 contig of 2052 bp in length.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 15221: contig of 15221 bp in length
* 15222 15321: gap of 100 bp
* 15322 30276: contig of 14955 bp in length
* 30277 30376: gap of 100 bp
* 30377 40753: contig of 10377 bp in length
* 40754 40853: gap of 100 bp
* 40854 53793: contig of 12940 bp in length
* 53794 53893: gap of 100 bp
* 53894 65011: contig of 11118 bp in length
* 65012 65111: gap of 100 bp
* 65112 75224: contig of 10113 bp in length
* 75225 75324: gap of 100 bp
* 75325 84814: contig of 9490 bp in length
* 84815 84914: gap of 100 bp
* 84915 96348: contig of 11434 bp in length
* 96349 96448: gap of 100 bp
```

	/note="assembly_fragment"	
misc_feature	.115022..123606	
	/note="assembly_fragment"	
misc_feature	.123707..131751	
	/note="assembly_fragment"	
misc_feature	.131852..138950	
	/note="assembly_fragment"	
misc_feature	.139051..144550	
	/note="assembly_fragment clone_end:SP6 vector_side:left"	
misc_feature	.144651..152953	
	/note="assembly_fragment"	
misc_feature	.153054..159224	
	/note="assembly_fragment"	
misc_feature	.159325..165736	
	/note="assembly_fragment"	
misc_feature	.165837..173107	
	/note="assembly_fragment"	
misc_feature	.173208..178488	
	/note="assembly_fragment"	
misc_feature	.178589..183587	
	/note="assembly_fragment"	
Query Match	1.0%; Score 32; DB 2; Length 224010;	
Best Local Similarity	100.0%; Pred. No. 0.00017;	
Matches 32:	Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 190	ttgaaaaaatTTTTTTTTTTTTTTTTTTT 221	
	TTTTTTTTTTTTTTTTTTTTTTTT	
Db 165693	TTGAAAAAATTTTTTTTTTTTTTTTTTTT 165724	
RESULT 14		
PFMAL13P3/C	PFMAL13P3 318221 bp DNA linear HTG 19-AUG-1999	
LOCUS	Plasmodium falciparum chromosome 13 strain 3D7, *** SEQUENCING IN	
DEFINITION	PROGRESS ***, in unordered pieces.	
ACCESSION	AL049184 .5 GI:5763803	
VERSION	AL049184.5	
KEYWORDS	HTG; HTGS_PHASE1.	
SOURCE	malaria parasite P. falciparum.	
ORGANISM	Plasmodium falciparum	
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
AUTHORS	1 (bases 1 to 318221) Bowman,S., Church,C., Harris,B., Harris,D., Lawson,D., Quall,M. and Barrel,B.	
TITLE	Direct Submission	
JOURNAL	Submitted (15-WAR-1999) P.falciparum Genome Sequencing Consortium.	

COMMENT
 On Aug 24, 1999 this sequence version replaced gi:5731882.
 For more information about this sequence or the Malaria Project, see <http://www.sanger.ac.uk/projects/P-falciparum>. IMPORTANT: This sequence is unfinished and does not necessarily represent the correct sequence. Work on the sequence is in progress and the release of this data is based on the understanding that the sequence may change as work continues. The sequence may be contaminated with foreign sequence from E.coli, yeast, vector, phage etc.
 Order of segments is not known; 800 n's separate segments.
 * NOTE: This is a 'working draft' sequence.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

FEATURES
 source 1. 318221
 /organism="Plasmodium falciparum"
 /strain="3D7"
 /db_xref="taxon:5833"
 /chromosome="13"

BASE COUNT 113610 a 29264 c 27321 g 118405 t 29621 others
 ORIGIN

```
Query Match      1.0%; Score 32; DB 2; Length 318221;
Best Local Similarity 100.0%; Pred. No. 0.00016;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 193 aaaaaattttttttttttttttttttttttttttgt 224
      |||||||
Db 101873 AAAAAAATTTTTTTTTTTTTTTTTTTTGT 101842

RESULT 15
LOCUS      CEZC334      35836 bp      DNA      linear      INV 24-JAN-2002
DEFINITION Caenorhabditis elegans cosmid ZC334, complete sequence.
ACCESSION  Z82082
VERSION    Z82082.1 GI:3218092
KEYWORDS   HTG.
SOURCE     Caenorhabditis elegans.
ORGANISM   Caenorhabditis elegans
            Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
            Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
REFERENCE  1 (sites)
AUTHORS    none.
TITLE      Genome sequence of the nematode C. elegans: a platform for
            investigating biology. The C. elegans Sequencing Consortium
            Science 282 (5396), 2012-2018 (1998)
            99069613
            The C. elegans Sequencing Consortium.
            2 (bases 1 to 35836)
            McLay,K.
            Direct Submission
            Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger
            Institute, Hinxton, Cambridge CB10 1SA, England and Department of
            Genetics, Washington University, St. Louis, MO 63110, USA. E-mail:
            jes@sanger.ac.uk or rv@nematode.wustl.edu
            On Jun 13, 1998 this sequence version replaced gi:1666043.
            Coding sequences below are predicted from computer analysis, using
            predictions from GeneFinder (P. Green, U. Washington), and other
            available information.
            Current sequence finishing criteria for the C. elegans genome
            sequencing consortium are that all bases are either sequenced
            unambiguously on both strands, or on a single strand with both a
            dye primer and dye terminator reaction, from distinct subclones.
            Exceptions are indicated by an explicit note.
            This sequence is the entire insert of clone ZC334. The start of
            this sequence (1..106) overlaps with the end of sequence AL132902.
            The end of this sequence (35735..35836) overlaps with the start of
            sequence AL023853.
            For a graphical representation of this sequence and its analysis
            see:- http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?
            name=ZC334
            Location/Qualifiers
              1..35836
                /organism="Caenorhabditis elegans"
                /db_xref="taxon:6239"
                /chromosome="I"
                /clone="ZC334"
            join(73..105,150..415,478..635,810..1122,1172..1762,
            2392..2767,4238..4361,4948..5183,5248..5356,5644..5831)
            /gene="ZC334.5"
            join(73..105,150..415,478..635,810..1122,1172..1762,
            2392..2767,4238..4361,4948..5183,5248..5356,5644..5831)
            /gene="ZC334.5"
            /note="predicted using GeneFinder
            similar to Leucine Rich Repeat (2 copies) (2 domains)
            CDNA EST EMBL:M75821 comes from this gene
            CDNA EST yk98f3.3 comes from this gene
            CDNA EST yk98f3.5 comes from this gene
            CDNA EST yk47a7.3 comes from this gene
            CDNA EST yk47a7.5 comes from this gene
```

```
CDNA EST yk284d4.3 comes from this gene
CDNA EST yk413d8.3 comes from this gene
CDNA EST yk450g9.3 comes from this gene
CDNA EST yk467f4.3 comes from this gene
CDNA EST yk491b5.3 comes from this gene
CDNA EST yk492c3.3 comes from this gene
CDNA EST yk284d4.5 comes from this gene
CDNA EST yk354e4.5 comes from this gene
CDNA EST yk413d8.5 comes from this gene
CDNA EST yk450g9.5 comes from this gene
CDNA EST yk492c3.5 comes from this gene
/codon_start=1
/protein_id="CAB04966.1"
/db_xref="GI:3881434"
/db_xref="SPTREMBL:O9XUI7"
/translation="MTSYSCSLAETGQLPWLRLSALCGLKFKYKTEFSPFVNL
DCLDISYNSVNLGVISNLQNLRLSGLGEFRQADLEELFTLKLNLVDMSLRLN
ALQTNIRVLSCNKLVPFLRLDLSGNKDEDELEELSHSLHKLTAICGCVKAKP
NCPIQSYTTYESTIFAIKHLRLRRRAELQSCFDLHSLKDYHEMGKLYDVERC
YLAVQSMKHLPSMP IFADRLMDAYFFYFADCLAEIVIHSHRVQDDISYICKET
KNVGRNLANTVNERICNATISLPSKGLSKDLKRLQELALNCMHALESTDOAAHDLA
RIILGSLNESNLVEMFSGIDVNPATKPVGLNFLNAVVRDDVETCKLVAEKVSRNTYN
KKECMOYINIGAMDVLISALGRFDSPLHTAILGIVRDLYVMAPYTLPKITFIQF
RTQVLKLTWTSORAYLDVLSILARLNHSDSSTECVFKKCADELKKOPFYLLATKIS
VDVDIYTINGIVONILQVSDVGITLWALLTIKIFAQKSPQSSQTIIRNSPNFFIQKP
RIILSEAVNMASENIDVPTQVMHYIHTFLATFLPLILADTLTLTADLTGHAASAAVF
QSEPSDDYFSTSDPSTSPRYPTDEADEELYALQKAAKPOFYPRPNNTGGYGGYRP
NPQPGYSYGVGINROARYPVMMYTLIOCPQORAKHLLATTYGDI PSHFLELVGD
EDWQRQLQVLLKVTQATFPYVFCGNFTGGSSDLFMHHSGQLRALNLCMSRNG"
join(16513..16668,16710..17585)
/gene="ZC334.4"
join(16513..16668,16710..17585)
/gene="ZC334.4"
/note="contains similarity to Pfam domain: PF01498
(Transposase), Score=208.2, E-value=4.1e-59, N=1"
/codon_start=1
/protein_id="CAB04965.2"
/db_xref="GI:18376576"
/db_xref="SPTREMBL:Q18285"
/translation="MVKSGCKNLSDVKKAI VAGFEQGIPTKMLAQIORSPTIWK
VIRKYOPEKSVALARISPRGRVTHRMDRLNLSAREDPHTATDIOMLISSPNPEVP
SKFTVRRRLQOAGLHGRKPKVKPFISKNNMARVAKAHLRWGRQSWAKHNSDESK
RFUYGSDGNSWNRVGVSRYSRKYQCTVXKHGGGVMMWGCFTSRRRPLKRIQSDP
FNQENILETTMRPWLQNVGRGFQDDNDPKHTSLHVSFSTFRRVHLLDWPQSP
DLNPIHLWELELRGGIRASNAADAKFNQLENAWKAIPMSVITHKLIDSPRRQCAVI
DANGYATKY"
complement(join(25665..25747,25791..26085))
/gene="ZC334.3"
complement(join(25665..25747,25791..26085))
/gene="ZC334.3"
/note="predicted using GeneFinder
CDNA EST yk636g11.3 comes from this gene
CDNA EST yk636g11.5 comes from this gene"
/codon_start=1
/protein_id="CAB04964.2"
/db_xref="GI:5824887"
/db_xref="SPTREMBL:Q9UIP6"
/translation="MTPPSHRYKNRNLNYDNSFVRMRSPFTFLLLLVPLALCHVFS
EPADLEUKSYQALEKSLKEMGLIRANQGPQKACGRSMKVKQLCAGCGTQNDLTI
KSCSTGYTDAGFTISACCPGCVF"
complement(join(29614..29708,30138..30405))
/gene="ZC334.2"
complement(join(29614..29708,30138..30405))
/gene="ZC334.2"
/note="predicted using GeneFinder
CDNA EST yk300g5.5 comes from this gene
CDNA EST yk440d9.5 comes from this gene"
/codon_start=1
/protein_id="CAB04963.1"
/db_xref="GI:3881431"
/db_xref="SPTREMBL:O9XUI8"
/translation="MSSHALVFLFLLLPVALGHFLSKPAPDPRTFNKRLAETLKE
LQDMLLIQAPREPVAQAQAKKTCGRSLLIKIQOLCHICTVHADDLHEITACMKGLTD
```



```

gene SOLINSCPPIPQTPPFVF"
complement(join(31630..31718,31887..32130))
/gene="ZC334.1"
CDS complement(join(31630..31718,31887..32130))
/gene="ZC334.1"
/note="predicted using GeneFinder
cDNA EST YK149h6.5 comes from this gene
cDNA EST YK222h5.3 comes from this gene
cDNA EST YK222h5.5 comes from this gene"
/codon_start=1
/protein_id="CAB04962.1"
/db_xref="GI:3881430"
/db_xref="SPTREMBL:Q9XU19"
/translation="MRALVALICLMALCHAAMLDELMQKEVQEFHMGNGMQEFMKN
GLNHHHGTGKAGLTCGNIIERVDKLCNGQCTRYNDALVIKSGRGSVDMEFNVACC
PTMKLFTH"
BASE COUNT 11311 a 6406 c 6543 g 11576 t
ORIGIN

Query Match      0.9%  Score 31;  DB 3;  Length 35836;
Best Local Similarity 100.0%;  Pred. No. 0.00064;
Matches 31;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

Qy 192 gaaataattttttttttttttttttttttttttttttt 222
|||||
Db 13392 GAAAAAATTTTTTTTTTTTTTTTTTTTTTTTGTG 13422

RESULT 16
AC027803/3
LOCUS AC027803.2 140798 bp DNA linear HTG 27-JUN-2000
DEFINITION Homo sapiens chromosome 17 clone RP11-323K14 map 17, LOW-PASS
SEQUENCE SAMPLING.
ACCESSION AC027803
VERSION AC027803.2 GI:8746876
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 140798)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-323K14
Unpublished
2 (bases 1 to 140798)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,E.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choquel,Y., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Glade,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Kleind,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Liu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrum,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu.X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 27, 2000 this sequence version replaced gi:7382645.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)

```

```

http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L9072
Center clone name: 323_K_14
-----

```

```

* NOTE: This record contains 164 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

1 732: contig of 732 bp in length
733 832: gap of 100 bp
833 1565: contig of 733 bp in length
1566 1665: gap of 100 bp
1666 2402: contig of 737 bp in length
2403 2502: gap of 100 bp
2503 3237: contig of 735 bp in length
3238 3337: gap of 100 bp
3338 4086: contig of 749 bp in length
4087 4186: gap of 100 bp
4187 4930: contig of 744 bp in length
4931 5030: gap of 100 bp
5031 5758: contig of 728 bp in length
5759 5858: gap of 100 bp
5859 6596: contig of 738 bp in length
6597 7436: contig of 740 bp in length
7437 7536: gap of 100 bp
7537 8247: contig of 711 bp in length
8248 8347: gap of 100 bp
8348 9093: contig of 746 bp in length
9094 9193: gap of 100 bp
9194 9940: contig of 747 bp in length
9941 10040: gap of 100 bp
10041 10764: contig of 724 bp in length
10765 10864: gap of 100 bp
10865 11579: contig of 715 bp in length
11580 11679: gap of 100 bp
11680 12419: contig of 740 bp in length
12420 12519: gap of 100 bp
12520 13253: contig of 734 bp in length
13254 13353: gap of 100 bp
13354 14097: contig of 744 bp in length
14098 14197: gap of 100 bp
14198 14950: contig of 753 bp in length
14951 15050: gap of 100 bp
15051 15791: contig of 741 bp in length
15792 15891: gap of 100 bp
15892 16614: contig of 723 bp in length
16615 16714: gap of 100 bp
16715 17413: contig of 699 bp in length
17414 17513: gap of 100 bp
17514 18224: contig of 711 bp in length
18225 18324: gap of 100 bp
18325 19078: contig of 754 bp in length
19079 19178: gap of 100 bp
19179 19885: contig of 707 bp in length
19886 19985: gap of 100 bp
19986 20713: contig of 728 bp in length
20714 20813: gap of 100 bp
20814 21548: contig of 735 bp in length
21549 21648: gap of 100 bp

```

```
* 21649 22386: contig of 738 bp in length
* 22387 22486: gap of 100 bp
* 22487 23222: contig of 736 bp in length
* 23223 23322: gap of 100 bp
* 23323 23422: contig of 720 bp in length
* 23423 24042: contig of 720 bp in length
* 24043 24142: gap of 100 bp
* 24143 24875: contig of 733 bp in length
* 24876 24975: gap of 100 bp
* 24976 25715: contig of 740 bp in length
* 25716 25815: gap of 100 bp
* 25816 26543: contig of 728 bp in length
* 26544 26643: gap of 100 bp
* 26644 27365: contig of 722 bp in length
* 27366 27465: gap of 100 bp
* 27466 28196: contig of 731 bp in length
* 28197 28296: gap of 100 bp
* 28297 29044: contig of 748 bp in length
* 29045 29144: gap of 100 bp
* 29145 29888: contig of 744 bp in length
* 29889 29988: gap of 100 bp
* 29989 30721: contig of 733 bp in length
* 30722 30821: gap of 100 bp
* 30822 31568: contig of 747 bp in length
* 31569 31668: gap of 100 bp
* 31669 32391: contig of 723 bp in length
* 32392 32491: gap of 100 bp
* 32492 33192: contig of 701 bp in length
* 33193 33292: gap of 100 bp
* 33293 34007: contig of 715 bp in length
* 34008 34107: gap of 100 bp
* 34108 34858: contig of 751 bp in length
* 34859 34958: gap of 100 bp
* 34959 35666: contig of 708 bp in length
* 35667 35766: gap of 100 bp
* 35767 36507: contig of 741 bp in length
* 36508 36607: gap of 100 bp
* 36608 37333: contig of 726 bp in length
* 37334 37433: gap of 100 bp
* 37434 38182: contig of 749 bp in length
* 38183 38282: gap of 100 bp
* 38283 39026: contig of 744 bp in length
* 39027 39126: gap of 100 bp
* 39127 39871: contig of 745 bp in length
* 39872 39971: gap of 100 bp
* 39972 40715: contig of 744 bp in length
* 40716 40815: gap of 100 bp
* 40816 41541: contig of 726 bp in length
* 41542 41641: gap of 100 bp
* 41643 42392: contig of 751 bp in length
* 42393 42492: gap of 100 bp
* 42493 43216: contig of 724 bp in length
* 43217 43316: gap of 100 bp
* 43317 44041: contig of 725 bp in length
* 44042 44141: gap of 100 bp
* 44142 44885: contig of 744 bp in length
* 44886 44985: gap of 100 bp
* 44986 45729: contig of 744 bp in length
* 45730 45829: gap of 100 bp
* 45830 46566: contig of 737 bp in length
* 46567 46666: gap of 100 bp
* 46667 47382: contig of 716 bp in length
* 47383 47482: gap of 100 bp
* 47483 48239: contig of 757 bp in length
* 48240 48339: gap of 100 bp
* 48340 49091: contig of 752 bp in length
* 49092 49191: gap of 100 bp
* 49192 49927: contig of 736 bp in length
* 49928 50027: gap of 100 bp
* 50028 50765: contig of 738 bp in length
* 50766 50865: gap of 100 bp
* 50866 51599: contig of 734 bp in length
* 51600 51699: gap of 100 bp
* 51700 52420: contig of 721 bp in length
```

```
* 52421 52520: gap of 100 bp
* 52521 53332: contig of 712 bp in length
* 53333 53332: gap of 100 bp
* 53333 54054: contig of 722 bp in length
* 54055 54154: gap of 100 bp
* 54155 54870: contig of 716 bp in length
* 54871 54970: gap of 100 bp
* 54971 55699: contig of 729 bp in length
* 55700 55799: gap of 100 bp
* 55800 56546: contig of 747 bp in length
* 56547 56646: gap of 100 bp
* 56647 57400: contig of 734 bp in length
```

Query Match 0.9%; Score 31; DB 2; Length 140798;
Best Local Similarity 100.0%; Pred. No. 0.00054;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 193 aaaaaattttttttttttttttttttgt 223
|||||
Db 22956 AAAAAAATTTTTTTTTTTTTTTTTTTGT 22926
```

RESULT 17

AL158193/c
LOCUS
DEFINITION
Human DNA sequence from clone RP11-400N12 on chromosome 13 Contains the end of the UBL3 (ubiquitin-like 3) gene, ESTs, STSS, GSSs and a CpG island, complete sequence.

ACCESSION
AL158193
VERSION
AL158193.13 GI:11182169
KEYWORDS
HTG; CpG island; ubiquitin; UBL3.
SOURCE
human.

ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 172896)
Wall, M.

REFERENCE
AUTHORS
TITLE
JOURNAL
Submitted (15-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

COMMENT

On Nov 16, 2000 this sequence version replaced gi:11137660.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em., EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/Projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 13, constructed by the Sanger Centre Chromosome 13 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr13>
This sequence is the entire insert of clone RP11-400N12 The true left end of clone RP11-174H18 is at 89870 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. RP11-400N12 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6.
FEATURES
source
1..172896
/organism="Homo sapiens"

```
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-400N12"
/clone_lib="RPC1-11.2"
690..1132
/note="MERC repeat: matches 173..602 of consensus"
1420..1559
/note="MERC repeat: matches 2..173 of consensus"
1743..2156
/note="match: GSS: Em:AQ151959"
complement(3569..4072)
/note="match: GSS: Em:AQ754516"
complement(3632..4071)
/note="match: GSS: Em:AQ703632"
9776..12321
/note="L1P47 repeat: matches 3601..6141 of consensus"
12645..16537
/note="L1P47 repeat: matches 18..3601 of consensus"
16538..17237
/note="L1P10 repeat: matches 5445..6164 of consensus"
17259..17390
/note="L1MBC repeat: matches 1897..2021 of consensus"
17394..17425
/note="16 copies 2 mer aa 84% conserved"
complement(19478..20162)
/note="match: GSS: Em:AQ350792"
complement(19942..20152)
/note="match: GSS: Em:AQ806482"
19955..20086
/note="L1R29 repeat: matches 23..149 of consensus"
20236..20512
/note="L1R29 repeat: matches 370..619 of consensus"
21147..26393
/note="L1HS repeat: matches 900..6146 of consensus"
26389..27137
/note="L1HS repeat: matches 32..776 of consensus"
26723..27139
/note="CpG island"
/evidence="not_experimental"
29280..29387
/note="54 copies 2 mer ta 60% conserved"
29459..29573
/note="23 copies 5 mer tatat 61% conserved"
29493..29590
/note="49 copies 2 mer ta 61% conserved"
29595..30568
/note="MER21B repeat: matches 14..751 of consensus"
complement(30597..30767)
32936..32961
/note="13 copies 2 mer ag 92% conserved"
complement(36176..36435)
/note="match: STS: Em:L18463"
36359..36402
/note="22 copies 2 mer ct 75% conserved"
36761..36838
/note="L1P2 repeat: matches 6077..6155 of consensus"
37666..38302
/note="L1P10 repeat: matches 5526..6165 of consensus"
39311..39444
/note="67 copies 2 mer aa 60% conserved"
39632..40222
/note="L1R10C repeat: matches 1..586 of consensus"
41331..41623
/note="MER2 repeat: matches 7..308 of consensus"
complement(41701..42375)
/note="match: GSS: Em:AQ746749"
41906..42375
/note="match: GSS: Em:AQ319140"
complement(41924..42375)
/note="match: GSS: Em:A2254542"
complement(41956..42375)
/note="match: GSS: Em:AQ489422"
complement(42087..42374)
/note="match: GSS: Em:AQ808740"
complement(42095..42374)
/note="match: GSS: Em:A2517696"
complement(42105..42375)
/note="match: GSS: Em:AQ817421"
complement(42115..42393)
/note="match: GSS: Em:B94003"
42147..42408
/note="match: GSS: Em:AQ627818"
complement(42154..42374)
/note="match: GSS: Em:AQ815465"
complement(42177..42375)
/note="match: GSS: Em:AQ617736"
complement(42221..42374)
/note="match: GSS: Em:B69794"
complement(42234..42377)
/note="match: GSS: Em:AQ021494"
complement(42236..42374)
/note="match: GSS: Em:AQ037473"
42391..42718
/note="MER2 repeat: matches 327..652 of consensus"
44721..51158
/note="L1P2 repeat: matches 8..6144 of consensus"
51186..52830
/note="L1P47 repeat: matches 4480..6138 of consensus"
53104..53442
/note="match: GSS: Em:AQ376651"
complement(56787..57241)
/note="match: GSS: Em:AQ126222"
61609..62001
/note="MER61B repeat: matches 1..396 of consensus"
complement(61809..62252)
/note="match: GSS: Em:AQ429742"
63783..64278
/note="match: GSS: Em:AQ614639"
67137..67325
/note="L1R33 repeat: matches 311..519 of consensus"
71184..71219
/note="18 copies 2 mer tt 91% conserved"
73436..73677
/note="match: GSS: Em:AQ724293"
73683..73841
/note="match: GSS: Em:AQ724293"
74516..74994
/note="match: STS: Em:G63026"
75968..76017
/note="25 copies 2 mer tt 74% conserved"
76175..76272
/note="49 copies 2 mer ta 61% conserved"
78172..78283
/note="56 copies 2 mer ta 65% conserved"
complement(78272..78706)
/note="match: GSS: Em:AQ355192"
78328..78836
/note="L1PBB repeat: matches -358..155 of consensus"
78830..81169
/note="L1P2 repeat: matches 3470..5821 of consensus"
81476..81801
/note="L1P2 repeat: matches 5821..6155 of consensus"
86340..86774
/note="match: GSS: Em:AQ713295"
86358..86957
/note="match: GSS: Em:AQ426776"
86889..86925
/note="Single clone region. Sequence from clone PCR only."
88375..88444
/note="35 copies 2 mer tt 67% conserved"
complement(89320..89853)
/note="match: GSS: Em:AQ763680"
complement(89397..89870)
/note="match: GSS: Em:AQ202624"
match: STS: Em:G61256"
```


REFERENCE
AUTHORS

2 (bases 1 to 19820)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Cooke, A., Choepel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P., DeArellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M., Fensterstor, J., Ferrisair, P., Fitzhugh, W., Forrest, C., Gage, D., Galagan, J., Gardyna, S., Glinde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Landers, T., Largocque, K., Lehoczyk, J., Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrim, J., Meneus, J., Milhova, T., Miranda, C., Mienga, V., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivat, T. M., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triglilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1970, Vol. 1, No. 2, pp. 1-10
3. The Importance of Parental Involvement	Journal of Educational Psychology	1975, Vol. 66, No. 3, pp. 1-10
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1980, Vol. 83, No. 4, pp. 1-10
5. The Role of the School in the Community	Journal of Educational Research	1985, Vol. 88, No. 5, pp. 1-10
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	1990, Vol. 93, No. 6, pp. 1-10
7. The Role of the Teacher in the Classroom	Journal of Educational Research	1995, Vol. 98, No. 7, pp. 1-10
8. The Impact of Technology on Education	Journal of Educational Technology	2000, Vol. 2, No. 8, pp. 1-10
9. The Importance of Parental Involvement	Journal of Educational Psychology	2005, Vol. 96, No. 9, pp. 1-10
10. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	2010, Vol. 113, No. 10, pp. 1-10
11. The Role of the School in the Community	Journal of Educational Research	2015, Vol. 118, No. 11, pp. 1-10
12. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2020, Vol. 123, No. 12, pp. 1-10

Direct Submission
Submitted (11-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 25, 2001 this sequence version replaced gi:7139742.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>
----- Genome Center

FEATURES SOURCE

```

clone_end:T7
vector_side:right"
/note="assembly_fragment"
9807. .9792
/note="assembly_fragment"
9893. .12319
/note="assembly_fragment"
12420. .14643
/note="assembly_fragment"
14744. .18502
/note="assembly_fragment"
18603. .22215
/note="assembly_fragment"
22316. .26466
/note="assembly_fragment"
26567. .31302
/note="assembly_fragment"
31403. .37782
/note="assembly_fragment"
37883. .42291
/note="assembly_fragment"
42392. .48984
/note="assembly_fragment"
49085. .56503
/note="assembly_fragment"
56604. .61901
/note="assembly_fragment"
clone_end:SP6
vector_side:right"
/note="assembly_fragment"
62002. .68691
/note="assembly_fragment"
68792. .76855
/note="assembly_fragment"
76956. .84821
/note="assembly_fragment"
84922. .96419
/note="assembly_fragment"
96520. .109449
/note="assembly_fragment"
109550. .125405
/note="assembly_fragment"
125506. .148312
/note="assembly_fragment"
148413. .173817
/note="assembly_fragment"
BASE COUNT 49665 a 37611 c 37301 g 46828 t 2412 others
ORIGIN

```

```

Query Match      0.9%; Score 31; DB 2; Length 173817;
Best Local Similarity 100.0%; Pred. No. 0.00053;
Matches 31; Conservative 0; Mismatches 0; Indels .0; Gaps 0;

QY 193 aaaaaaaattttttttttttttttttttttt 223
|||||
Db 166689 AAAAAAATTTTTTTTTTTTTTTTTTTTGT 166659
|||||

```

RESULT	19
AC023315/c	
LOCUS	
DEFINITION	AC023315 198820 bp DNA linear HTG 25-JUN-2001 Homo sapiens clone RP11-333J10, WORKING DRAFT SEQUENCE, 7 unordered pieces.
ACCESSION	AC023315
VERSION	AC023315.3 GI:13123390
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE	human.
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 198820)
REFERENCE	Birren, B., Linton, J., Nusbaum, C. and Lander, E.
AUTHORS	Homo sapiens, clone RP11-333J10
TITLE	Unpublished
JOURNAL	

misc_feature /db_xref="taxon:9606"
/clone="RP11-333J10"
/clone_lib="RPCI-11 Human Male BAC"
1. .5936
clone_end:SP6
vector_side:left
6037. .17238
/note="assembly_fragment"
misc_feature 17339. .30763
/note="assembly_fragment"
misc_feature 30864. .50814
/note="assembly_fragment"
misc_feature 50915. .73714
/note="assembly_fragment"
misc_feature 73815. .117169
/note="assembly_fragment"
misc_feature 117270. .198820
/note="assembly_fragment"
clone_end:T7
vector_side:right
BASE COUNT 54514 a 44459 c 43520 g 55727 t 600 others
ORIGIN

Query Match 0.9%; Score 31; DB 2; Length 198820;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 193 aaaaaattttttttttttttttttttttgt 223
|||||
Db 18920 AAAAAAAAAATTTTttttttttttttttttgt 18890

RESULT 20
AC079869/c AL671090 198932 bp DNA linear HTG 08-FEB-2002
LOCUS Mus musculus chromosome 4 clone RP23-154A10, *** SEQUENCING IN
DEFINITION PROGRESS ***, in unordered pieces.
ACCESSION AL671090
VERSION AL671090.4 GI:18643845
KEYWORDS HTG; HTGS-PHASE1.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (sites)
Plumb.B.
Direct Submission
Submitted (07-FEB-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Feb 10, 2002 this sequence version replaced gi.18491462.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: <http://www.sanger.ac.uk>
Contact: humquerry@sanger.ac.uk
----- Project Information
Center project name: BM154A10
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 95% of reads
Chemistry: Dye-terminator; 4% of reads
Consensus quality: 197921 bases at least Q40
Consensus quality: 198121 bases at least Q30
Consensus quality: 198350 bases at least Q20
Insert size: 198732; sum-of-contigs
Insert size: 194340; 2.7% error; agarose-fp
Quality coverage: 10.28x in Q20 bases; sum-of-contigs Quality
coverage: 10.51x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
source Location/Qualifiers
1. .198932
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="4"
/clone="RP23-154A10"
/clone_lib="RPCI-23"
1. .65901
/note="assembly_fragment:00047"
66002. .192023
/note="assembly_fragment:02231"
fragment_chain:1"
192124. .198932
/note="assembly_fragment:01866"
fragment_chain:1
clone_end:T7
vector_side:right
BASE COUNT 66075 a 35350 c 35808 g 61499 t 200 others
ORIGIN

Query Match 0.9%; Score 31; DB 2; Length 198932;
Best Local Similarity 100.0%; Pred. No. 0.00052;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1519 gagaagagaaggaaggaaggaaggaaggg 1549
|||||
Db 24229 GAGAAGGAGAGGGGAGGAGGAGGAGG 24199

RESULT 21
AC079869/c AC079869 243675 bp DNA linear ROD 11-JUL-2001
LOCUS Mus musculus 9 BAC RP23-49M18 (Roswell Park Cancer Institute Mouse
DEFINITION BAC Library) complete sequence.
ACCESSION AC079869
VERSION AC079869.22 GI:14670076
KEYWORDS HTG.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 243675)
Metzker, M.L., Lewis, L.R., Hume, J., Edwards, C., Harris, C.,
Dederich, D., Thomas, S., Okwuonu, G., Carlock, C., Garner, T.,
Addison, S., Pace, A., Williams, G., Bonnin, D., Brooks, A., Brown, J.,
Buha, C., Bunac, C., Burkett, C., Chacko, J., Chen, G., Chen, Z.,
Cox, C., Davis, C., Delgado, O., Ding, Y., Dugan-Rocha, S.,
Fernandez, C., Ferraguto, D., Forcum-Tansey, J., Gill, R.,
Gorell, J.H., Gunaratne, P., Haller, G., Hernandez, J., Hogues, M.,
Hosak, H., Hou, X., Hubert, J., Jackson, L., Jia, Y., Kelly, J., Kelly, S.,
Kovar, C., Liu, J., Liu, W., Loulsegad, H., Lozado, R.J., Martin, R.,
Massey, E., McLeod, M.P., Mei, G., Moore, S., Morgan, M., Morris, S.,
Neal, D., Nelson, A., Nguyen, R., Nguyen, N., Ogih, M., Parish, B.,
Perez, L., Reiter, D., Say, J., Shen, H., Vasquez, L., Watlington, S.,
Williamson, A., Wrensford, G., Zhou, X., Bouck, J., Hodgson, A.,
Muzny, D.M., Rives, M., Scherer, S., Sodergren, E., Weinstock, G.,
Worley, K. and Gibbs, R.
Direct Submission
Unpublished
REFERENCE 2 (bases 1 to 243675)
Worley, K.C.
Direct Submission
Submitted (15-SEP-2000) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 243675)
Worley, K.C.
Direct Submission
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE

JOURNAL

Submitted (11-JUL-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 11, 2001 this sequence version replaced gi:13899377. Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality does not meet this standard, it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui Zhang.

Exon/Intron boundaries of identified genes were chosen if there were canonical splice junctions that maintained sequence continuity across the splice junctions.

FEATURES

source

1. .243675

/organism="Mus musculus"

/db_xref="taxon:10090"

/chromosome="9"

/clone="RP23-49M18"

20. .105

/rpt_family="CT-rich"

complement(575..721)

/rpt_family="Bl_MM"

1231..1259

/rpt_family="AT-rich"

1591..1735

/rpt_family="B2"

1786..1841

/rpt_family="B2"

1942..2027

/rpt_family="Bl-F"

2123..2230

/rpt_family="Bl_MM"

2285..2376

/rpt_family="L1MA7"

2407..2578

/rpt_family="Bl_MM"

2708..2732

/rpt_family="AT-rich"

3444..3568

/rpt_family="Bl-F"

3928..4076

/rpt_family="Bl_MM"

4091..4196

/rpt_family="RSINE1"

complement(4301..4496)

/rpt_family="B2"

complement(4497..4702)

/rpt_family="B3"

complement(4954..5065)

/rpt_family="PB1D10"

complement(5167..5346)

/rpt_family="B4"

6052..6243

/rpt_family="MTA"

complement(6394..6573)

/rpt_family="B2"

6751..6915

/rpt_family="B3"

7471..7615

/rpt_family="Bl_MM"

8160..8329

/rpt_family="B2"

8385..8476

/rpt_family="B2"

11055..11342

/rpt_family="B4A"

repeat_region

/rpt_family="L1"

repeat_region

/rpt_family="B2"

repeat_region

/rpt_family="Bl_MM"

repeat_region

/rpt_family="(CAAAA)n"

repeat_region

/rpt_family="Lx2"

repeat_region

/rpt_family="B2"

repeat_region

15288..15436

/rpt_family="Bl-F"

complement(15444..15708)

repeat_region

/rpt_family="L1MB8"

repeat_region

16059..16118

/rpt_family="AT-rich"

complement(17083..17442)

repeat_region

/rpt_family="Lx6"

repeat_region

complement(17445..17538)

repeat_region

complement(17474..17573)

repeat_region

/rpt_family="L1"

repeat_region

/rpt_family="PB1D9"

repeat_region

17774..17797

/rpt_family="(CAA)n"

repeat_region

17887..18105

/rpt_family="URRIA"

repeat_region

18984..19004

/rpt_family="AT-rich"

repeat_region

complement(19009..19124)

repeat_region

/rpt_family="Bl-F"

repeat_region

19343..19369

/rpt_family="(GA)n"

repeat_region

complement(19404..19499)

repeat_region

/rpt_family="B2"

repeat_region

20429..20470

/rpt_family="(TTGG)n"

repeat_region

complement(20471..20604)

repeat_region

/rpt_family="Bl_MM"

repeat_region

21476..21587

/rpt_family="B3"

repeat_region

22296..22441

/rpt_family="RSINE1"

repeat_region

23898..24032

/rpt_family="Bl_MM"

repeat_region

24353..24403

/rpt_family="(TTTA)n"

repeat_region

complement(24422..24532)

repeat_region

/rpt_family="Bl_MM"

repeat_region

24959..25083

/rpt_family="Tigger4(Zombi)"

complement(25171..25477)

/rpt_family="MTC"

25484..25570

/rpt_family="Tigger4(Zombi)"

complement(25975..26126)

/rpt_family="Bl_MM"

26504..26584

/rpt_family="(CA)n"

27439..27493

/rpt_family="(CA)n"

28100..28264

/rpt_family="(GAAA)n"

28553..28622

/rpt_family="(TATC)n"

complement(28662..28786)

/rpt_family="RSINE1"

28787..28989

/rpt_family="B2"

complement(29170..29378)

repeat_region

repeat_region /rpt_family="B2"
complement(29457..29604)
repeat_region /rpt_family="B1-F"
complement(30224..30344)
repeat_region /rpt_family="P81"
complement(30594..30753)
repeat_region /rpt_family="RSINEL"
30762..30785
repeat_region /rpt_family="AT_rich"
31123..31147
repeat_region /rpt_family="(TG)n"
complement(31302..31433)
repeat_region /rpt_family="B1-F"
31444..31503
repeat_region /rpt_family="(TA)n"
complement(31553..31665)
repeat_region /rpt_family="RSINEL"

Query Match 0.9%; Score 31; DB 10; Length 243675;
Best Local Similarity 100.0%; Pred. No. 0.00051;
Matches 31; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 191 tgaaaaaatTTTTTTTTTTTTTTTTTTT 221
Db 137978 TGAAAAAATTTTTTTTTTTTTTTTTT 137948

RESULT 22

AX277981 AX277981 8801 bp DNA linear PAT 01-NOV-2001
LOCUS Sequence 144 from Patent WO0177375.
DEFINITION AX277981
ACCESSION AX277981
VERSION AX277981.1 GI:16605053
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (sites)
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with gene regulation
JOURNAL Patent: WO 0177375-A 144 18-OCT-2001;
Epigenomics AG (DE)

FEATURES
source
1..8801
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2344 a 366 c 2057 g 4034 t
ORIGIN

Query Match 0.9%; Score 30; DB 6; Length 8801;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 gaataaaatTTTTTTTTTTTTTTTTTT 221
Db 971 GAAAAAATTTTTTTTTTTTTTTTTT 1000

RESULT 23

AX323672 AX323672 8801 bp DNA linear PAT 07-JAN-2002
LOCUS Sequence 160 from Patent WO0192565.
DEFINITION AX323672
ACCESSION AX323672
VERSION AX323672.1 GI:18094420
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (sites)
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with dna transcription

JOURNAL Patent: WO 0192565-A 160 06-DEC-2001;
Epigenomics AG (DE)
FEATURES Location/Qualifiers
source
1..8801
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2344 a 366 c 2057 g 4034 t
ORIGIN

Query Match 0.9%; Score 30; DB 6; Length 8801;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 gaataaaatTTTTTTTTTTTTTTTTTT 221
Db 971 GAAAAAATTTTTTTTTTTTTTTTTT 1000

RESULT 24

AX346643 AX346643 8801 bp DNA linear PAT 01-FEB-2002
LOCUS Sequence 1714 from Patent WO0200928.
DEFINITION AX346643
ACCESSION AX346643
VERSION AX346643.1 GI:18494529
KEYWORDS
SOURCE synthetic construct.
ORGANISM artificial sequence.
REFERENCE 1 (sites)
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 1714 03-JAN-2002;
Epigenomics AG (DE)

FEATURES
source
1..8801
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

BASE COUNT 2344 a 366 c 2057 g 4034 t
ORIGIN

Query Match 0.9%; Score 30; DB 6; Length 8801;
Best Local Similarity 100.0%; Pred. No. 0.0023;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 192 gaataaaatTTTTTTTTTTTTTTTTTT 221
Db 971 GAAAAAATTTTTTTTTTTTTTTTTT 1000

RESULT 25

AX011712/c HSA011712 16689 bp DNA linear PRI 27-APR-1999
LOCUS Homo sapiens TNNT1 gene, exons 1-11 (and joined CDS).
DEFINITION AJ011712
ACCESSION AJ011712
VERSION AJ011712.1 GI:4056561
KEYWORDS slow skeletal; TNNT1 gene; troponin T.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 16689)
AUTHORS Barton, P.J., Cullen, M.E., Townsend, P.J., Brand, N.J., Mullen, A.J., Norman, D.A., Bhavsar, P.K. and Yacoub, M.H.
TITLE Close physical linkage of human troponin genes: organization, sequence, and expression of the locus encoding cardiac troponin I and slow skeletal troponin T
JOURNAL Genomics 57 (1), 102-109 (1999)
MEDLINE 99208666
REFERENCE 2 (bases 1 to 16689)

AUTHORS Barton, P.J.R.
TITLE Direct Submission
JOURNAL Submitted (15-SEP-1998) Barton P.J.R., Cardiothoracic Surgery, National Heart & Lung Institute, Dovehouse Street, London SW3 6JB, United Kingdom

REMARK Revised on 20-DEC-1998
COMMENT Related sequences : X90780 (cardiac troponin I gene, directly upstream)
AJ011713 (TNNT1 gene, exons 12-14).

FEATURES
source 1..16689 Location/Qualifiers
`/organism="Homo sapiens"`
`/db_xref="taxon:9606"`
`/chromosome="19"`
`/map="q13.4"`
`/clone="DMPC-HFF#1-1495-E10"`
1..2580
`/gene="TNNT1"`
`/note="promoter region"`
join(1..16689,AJ011713.1:1..1846)
`/gene="TNNT1"`
2552..2557
`/gene="TNNT1"`
`/note="putative TATA box"`
join(2581..16689,AJ011713.1:1..1846)
`/gene="TNNT1"`
2581..2626
`/gene="TNNT1"`
`/number=1`
join(2581..2626,4587..4629,4734..4747,5048..5074,6189..6210,9829..9892,10447..10563,10789..13792,14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,AJ011713.1:1715..1760)
`/note="alternatively spliced mRNA"`
`/product="troponin T, slow skeletal muscle"`
join(2581..2626,4587..4629,4734..4747,5048..5074,5289..5321,6189..6210,9829..9892,10447..10563,10789..13792,14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,AJ011713.1:1715..1760)
`/gene="TNNT1"`
`/note="alternatively spliced mRNA"`
`/product="troponin T, skeletal muscle"`
join(2581..2626,4587..4629,4734..4747,5048..5074,5289..5321,6189..6210,9829..9892,10447..10563,10789..13792,14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,AJ011713.1:1715..1760)
`/gene="TNNT1"`
`/note="alternatively spliced mRNA"`
`/product="troponin T, skeletal and cardiac muscle"`
join(2581..2626,4587..4597)
`/gene="TNNT1"`
4587..4629
`/gene="TNNT1"`
`/number=2`
join(4598..4629,4734..4747,5048..5074,5289..5321,6189..6210,9829..9892,10447..10563,10789..13792,14577..14686,AJ011713.1:471..609,AJ011713.1:748..788,AJ011713.1:1715..1760)
`/gene="TNNT1"`
`/note="alternatively spliced isoform"`
`/codon_start=1`
`/product="slow skeletal muscle troponin T"`
`/protein_id="CAA09752.1"`
`/db_xref="GI:4056564"`
`/translation="MSDTEQEYEEQPEEAEAEPEEPVPAEPEERPKPSR
PVVPLIPPKIPEGERVDFDDIHRKMEKDLLEQTLIDVHFQKKEEELVALKER
TERRSRAEQRTEKERERQAKLAEEKMKKEEAKKRAEDDAKKKKVLSNNGAH
FGYLVKAEQKRGKQRTGRENKVRILSERKKPLDIDYNGEQLREKAQELSDWIHQLE
AREKAQELSDWIHQLESEKFDLMKLLKQKQKYEINVLNRIISHAOKFRKGRVGRWK"`

CDS
join(4598..4629,4734..4747,5048..5074,5289..5321,6189..6210,9829..9892,10447..10563,10789..13792,14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,AJ011713.1:1715..1760)
`/gene="TNNT1"`
`/note="alternatively spliced isoform"`
`/codon_start=1`
`/product="slow skeletal muscle troponin T"`
`/protein_id="CAA09751.1"`
`/db_xref="GI:4056563"`
`/translation="MSDTEQEYEEQPEEAEAEPEEPVPAEPEERPKPSR
PVVPLIPPKIPEGERVDFDDIHRKMEKDLLEQTLIDVHFQKKEEELVALKER
TERRSRAEQRTEKERERQAKLAEEKMKKEEAKKRAEDDAKKKKVLSNNGAH
FGYLVKAEQKRGKQRTGRENKVRILSERKKPLDIDYNGEQLREKAQELSDWIHQLE
SEKFDLMKLLKQKQKYEINVLNRIISHAOKFRKGRVGRWK"`
join(4598..4629,4734..4747,5048..5074,5289..5321,6189..6210,9829..9892,10447..10563,10789..13792,14577..14686,AJ011713.1:519..609,AJ011713.1:748..788,AJ011713.1:1715..1760)
`/gene="TNNT1"`
`/note="alternatively spliced isoform"`
`/codon_start=1`
`/product="slow skeletal muscle troponin T"`
`/protein_id="CAA09750.1"`
`/db_xref="GI:4056562"`
`/translation="MSDTEQEYEEQPEEAEAEPEEPVPAEPEERPKPSR
PEGERVDFDDIHRKMEKDLLEQTLIDVHFQKKEEELVALKER
TERRSRAEQRTEKERERQAKLAEEKMKKEEAKKRAEDDAKKKKVLSNNGAH
FGYLVKAEQKRGKQRTGRENKVRILSERKKPLDIDYNGEQLREKAQELSDWIHQLE
QKQKYEINVLNRIISHAOKFRKGRVGRWK"`
4734..4747
`/gene="TNNT1"`
`/number=3`
5048..5074
`/gene="TNNT1"`
`/number=4`
5289..5321
`/gene="TNNT1"`
`/note="alternative exon"`
`/number=5`
6189..6210
`/gene="TNNT1"`
`/number=6`
9829..9892
`/gene="TNNT1"`
`/number=7`
10447..10563
`/gene="TNNT1"`
`/number=8`
10789..10866
`/gene="TNNT1"`
`/number=9`
13679..13792
`/gene="TNNT1"`
`/number=10`
14577..14686
`/gene="TNNT1"`
`/number=11`
14687..>16689
`/gene="TNNT1"`
`/note="approx. 200 bp gap with AJ011713"`
`/number=11`

BASE COUNT 3888 a 4422 g 4743 g 3615 t 21 others
ORIGIN

Query Match 0.9%; Score 30; DB 9; Length 16689;
Best Local Similarity 100.0%; Pred. No. 0.0021;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1526 aqaaggggagagagagagagaggggaggg 1555
|||||
Db 6448 AGAGGGGAAGGAGGAGGAGGAGG 6419

RESULT 26

AC093149/c

LOCUS

DEFINITION

AC093149

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

JOURNAL

AC093149 Homo sapiens chromosome 7 clone RP11-382M23, complete sequence.
AC093149 HTG.
AC093149.1 GI:15148147

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 49816)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,
Clendenning, J., Ivey, R.G. and Haugen, E.D.

Direct Submission
Unpublished
2 (bases 1 to 49816)
Kaul, R.K., Olson, M.V., Raymond, C., Clendenning, J., Ivey, R.G. and
Haugen, E.D.

Submitted (11-AUG-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

3 (bases 1 to 49816)
Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Raymond, C.,
Clendenning, J., Ivey, R.G. and Haugen, E.D.

Direct Submission
Submitted (15-AUG-2001) Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA

Center: University of Washington Genome Center
Center Code: UWGC
Web site: <http://www.genome.washington.edu>
Contact: uwgctg@seu.washington.edu

Project Information
Center project name: chr-7
Center clone name: (djs734g)

Summary Statistics
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319
Consensus quality: 56066 bases at least Q40
Consensus quality: 57719 bases at least Q30
Consensus quality: 59069 bases at least Q20

Insert size: 61550; 25.6% error; agarose-fp
Insert size: 49816; sum-of-contigs
Quality coverage: 6.1x in Q20 bases; agarose-fp
Quality coverage: 7.6x in Q20 bases; sum-of-contigs

Overlapping Sequences:
5': RP11-702D16 (UWGC:djs709) AC083866, 46120-bp clone
overlap, of which only 4238 bp is included in this submission.

3': RP11-305M3 (UWGC:djs195) AC009244, 87331-bp clone overlap,
of which only 1340 bp is included in this submission.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	SeqDerMap	FrngPrnt	BglII
ECORI		HindIII				
9086	9202	949	959	3847	3886	
2035	2027	10444	10529	5756	5769	
848	855	11998	11988	1201	1189	
509	<800	875	874	5289	5293	
20482	20457	282	<800	2029	2044	
9908	9895	5873	5813	67	<800	
7709	7695	3805	3859	773	<800	
4788	4722	612	<800	1174	1189	
3787	3785	1094	1080	7149	7143	
		6289	6242	4364	4350	
		16931	16909	9	<800	
				6683	6696	
				3125	3073	
				3099	3073	
				2202	2220	
				6473	6441	
				3845	3886	
				2067	2044	

FEATURES

source

Location/Qualifiers
1..49816
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/clone="RP11-382M23"
/clone_lib="RPCI human BAC library 11"
29955..29967
/note="Single subclone region"
complement(44062..44179)
/standard_name="sWS31664"
BASE COUNT 14395 a 9205 c 10208 g 16008 t
ORIGIN

misc_feature

STS

Query Match

Best Local Similarity 100.0%; Score 30; DB 9; Length 49816;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
Theodore,J., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemбек,L., Zimmer,A. and Zody,M.

TITLE
JOURNAL
COMMENT

Submitted (07-MAY-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L13222

Center Clone name: 960_P_18

* NOTE: This record contains 80 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 720: contig of 720 bp in length
* 721 820: gap of 100 bp
* 821 1575: contig of 755 bp in length
* 1576 1675: gap of 100 bp
* 1676 2407: contig of 732 bp in length
* 2408 2507: gap of 100 bp
* 2508 3249: contig of 742 bp in length
* 3250 3349: gap of 100 bp
* 3350 4100: contig of 751 bp in length
* 4101 4200: gap of 100 bp
* 4201 4948: contig of 748 bp in length
* 4949 5048: gap of 100 bp
* 5049 5792: contig of 744 bp in length
* 5793 5892: gap of 100 bp
* 5893 6629: contig of 737 bp in length
* 6630 6729: gap of 100 bp
* 6730 7494: contig of 765 bp in length
* 7495 7594: gap of 100 bp
* 7595 8292: contig of 698 bp in length
* 8293 8392: gap of 100 bp
* 8393 9137: contig of 745 bp in length
* 9138 9237: gap of 100 bp
* 9238 9959: contig of 722 bp in length
* 9960 10059: gap of 100 bp
* 10060 10806: contig of 747 bp in length
* 10807 10906: gap of 100 bp
* 10907 11649: contig of 743 bp in length
* 11650 11749: gap of 100 bp
* 11750 12460: contig of 711 bp in length
* 12461 12560: gap of 100 bp
* 12561 13324: contig of 764 bp in length
* 13325 13424: gap of 100 bp
* 13425 14169: contig of 745 bp in length
* 14170 14269: gap of 100 bp
* 14270 15043: contig of 774 bp in length
* 15044 15143: gap of 100 bp
* 15144 15919: contig of 776 bp in length
* 15920 16019: gap of 100 bp
* 16020 16737: contig of 718 bp in length
* 16738 16837: gap of 100 bp
* 16838 17518: contig of 681 bp in length
* 17519 17618: gap of 100 bp
* 17619 18341: contig of 723 bp in length

* 18342 18441: gap of 100 bp
* 18442 19167: contig of 726 bp in length
* 19168 19267: gap of 100 bp
* 19268 19999: contig of 732 bp in length
* 20000 20099: gap of 100 bp
* 20100 20835: contig of 736 bp in length
* 20836 20935: gap of 100 bp
* 20936 21690: contig of 755 bp in length
* 21691 21790: gap of 100 bp
* 21791 22527: contig of 737 bp in length
* 22528 22627: gap of 100 bp
* 22628 23387: contig of 760 bp in length
* 23388 23487: gap of 100 bp
* 23488 24234: contig of 747 bp in length
* 24235 24334: gap of 100 bp
* 24335 25086: contig of 752 bp in length
* 25087 25186: gap of 100 bp
* 25187 25934: contig of 748 bp in length
* 25935 26034: gap of 100 bp
* 26035 26757: contig of 723 bp in length
* 26758 26857: gap of 100 bp
* 26858 27595: contig of 738 bp in length
* 27596 27695: gap of 100 bp
* 27696 28428: contig of 733 bp in length
* 28429 28528: gap of 100 bp
* 28529 29287: contig of 759 bp in length
* 29288 29387: gap of 100 bp
* 29388 30133: contig of 746 bp in length
* 30134 30233: gap of 100 bp
* 30234 30954: contig of 721 bp in length
* 30955 31054: gap of 100 bp
* 31055 31795: contig of 741 bp in length
* 31796 31895: gap of 100 bp
* 31896 32644: contig of 749 bp in length
* 32645 32744: gap of 100 bp
* 32745 33516: contig of 772 bp in length
* 33517 33616: gap of 100 bp
* 33617 34369: contig of 753 bp in length
* 34370 34469: gap of 100 bp
* 34470 35200: contig of 731 bp in length
* 35201 35300: gap of 100 bp
* 35301 36048: contig of 748 bp in length
* 36049 36148: gap of 100 bp
* 36149 36894: contig of 746 bp in length
* 36895 36994: gap of 100 bp
* 36995 37729: contig of 735 bp in length
* 37730 37829: gap of 100 bp
* 37830 38565: contig of 736 bp in length
* 38566 38665: gap of 100 bp
* 38666 39412: contig of 747 bp in length
* 39413 39512: gap of 100 bp
* 39513 40273: contig of 761 bp in length
* 40274 40373: gap of 100 bp
* 40374 41112: contig of 739 bp in length
* 41113 41212: gap of 100 bp
* 41213 41969: contig of 757 bp in length
* 41970 42069: gap of 100 bp
* 42070 42819: contig of 750 bp in length
* 42820 42919: gap of 100 bp
* 42920 43671: contig of 752 bp in length
* 43672 43771: gap of 100 bp
* 43772 44509: contig of 738 bp in length
* 44510 44609: gap of 100 bp
* 44610 45346: contig of 737 bp in length
* 45347 45446: gap of 100 bp
* 45447 46213: contig of 767 bp in length
* 46214 46313: gap of 100 bp
* 46314 47076: contig of 763 bp in length
* 47077 47176: gap of 100 bp
* 47177 47929: contig of 753 bp in length
* 47930 48029: gap of 100 bp
* 48030 48784: contig of 755 bp in length
* 48785 48884: gap of 100 bp

```

* 48885 49635: contig of 751 bp in length
* 49636 49735: gap of 100 bp
* 49736 50474: contig of 739 bp in length
* 50475 50574: gap of 100 bp
* 50575 51317: contig of 743 bp in length
* 51318 51417: gap of 100 bp
* 51418 52169: contig of 752 bp in length
* 52170 52269: gap of 100 bp
* 52270 53008: contig of 739 bp in length
* 53009 53108: gap of 100 bp
* 53109 53861: contig of 753 bp in length
* 53862 53961: gap of 100 bp
* 53962 54724: contig of 763 bp in length
* 54725 54824: gap of 100 bp
* 54825 55571: contig of 747 bp in length
* 55572 55671: gap of 100 bp
* 55672 56448: contig of 777 bp in length
* 56449 56548: gap of 100 bp
* 56549 57289: contig of 741 bp in length
* 57290 57389: gap of 100 bp
* 57390 58100: contig of 711 bp in length
* 58101 58200: gap of 100 bp

```

```

Query Match          0.9%: Score 30; DB 2; Length 67462;
Best Local Similarity 100.0%; Pred. No. 0.0018;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 193 aaaaaattttttttttttttttttttttttt 222
|||||
Db 21115 AAAAAATTTTTTTTTTTTTTTTTTTTG 21086

```

```

RESULT 30
AC109316
LOCUS Homo sapiens chromosome 8 clone RP11-96H13 map 8, LOW-PASS SEQUENCE
DEFINITION AC109316
AC109316.1 GI:18482245
VERSION -HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 73858)
Birren,B., Linton,L., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 8, clone RP11-96H13
Unpublished
2 (bases 1 to 73858)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepe,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M.,
McEwan,P., McKernan,K., Meldrum,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Raymond,C.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomand,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye.W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (03-FEB-2002) Whitehead Institute/MIT Center for Genome

```

COMMENT

```

Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23111
Center clone name: 96_H13
-----

```

```

* NOTE: This record contains 90 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

```

```

* 1 713: contig of 713 bp in length
* 714 813: gap of 100 bp
* 814 1531: contig of 718 bp in length
* 1532 1631: gap of 100 bp
* 1632 2360: contig of 729 bp in length
* 2361 2460: gap of 100 bp
* 2461 3148: contig of 688 bp in length
* 3149 3248: gap of 100 bp
* 3249 3939: contig of 691 bp in length
* 3940 4039: gap of 100 bp
* 4040 4757: contig of 718 bp in length
* 4758 4857: gap of 100 bp
* 4858 5576: contig of 719 bp in length
* 5577 5676: gap of 100 bp
* 5677 6390: contig of 714 bp in length
* 6391 6490: gap of 100 bp
* 6491 7212: contig of 722 bp in length
* 7213 7312: gap of 100 bp
* 7313 8025: contig of 713 bp in length
* 8026 8125: gap of 100 bp
* 8126 8865: contig of 740 bp in length
* 8866 8965: gap of 100 bp
* 8966 9679: contig of 714 bp in length
* 9680 9779: gap of 100 bp
* 9780 10518: contig of 739 bp in length
* 10519 10618: gap of 100 bp
* 10619 11349: contig of 731 bp in length
* 11350 11449: gap of 100 bp
* 11450 12174: contig of 725 bp in length
* 12175 12274: gap of 100 bp
* 12275 12987: contig of 713 bp in length
* 12988 13087: gap of 100 bp
* 13088 13779: contig of 692 bp in length
* 13780 13879: gap of 100 bp
* 13880 14593: contig of 714 bp in length
* 14594 14693: gap of 100 bp
* 14694 15420: contig of 727 bp in length
* 15421 15520: gap of 100 bp
* 15521 16256: contig of 736 bp in length
* 16257 16356: gap of 100 bp
* 16357 17093: contig of 737 bp in length
* 17094 17193: gap of 100 bp
* 17194 17920: contig of 727 bp in length
* 17921 18020: gap of 100 bp
* 18021 18730: contig of 710 bp in length
* 18731 18830: gap of 100 bp
* 18831 19560: contig of 730 bp in length
* 19561 19660: gap of 100 bp
* 19661 20394: contig of 734 bp in length

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 14, 2002, 17:02:26 ; Search time 72.29 seconds
(without alignments)
11338.761 Million cell updates/sec

Title: US-09-667-130-1

Perfect score: 3337

Sequence: 1 gaattccggtaaagtaacaa.....tcttataaataaataattc 3337

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 383533 seqs, 122816752 residues

Word size : 0

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*

5: /cgn2_6/ptodata/1/ina/PCBUS_COMB.seq.*

6: /cgn2_6/ptodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3337	100.0	3337	1 US-08-072-610-1	Sequence 1, Appli
2	3337	100.0	3337	2 US-08-719-822B-1	Sequence 1, Appli
3	3337	100.0	3337	4 US-09-092-458-1	Sequence 1, Appli
4	29	0.9	32	5 PCT-US92-10792-1	Sequence 16, Appli
5	28	0.8	90	4 US-09-065-058-16	Sequence 7, Appli
6	27	0.8	50	2 US-08-828-007-7	Sequence 4, Appli
7	27	0.8	3138	1 US-07-867-106-4	Sequence 1, Appli
8	26	0.8	76	4 US-09-244-794A-1	Sequence 1, Appli
9	26	0.8	76	4 US-09-247-190-1	Sequence 1, Appli
10	26	0.8	596	1 US-08-150-331-45	Sequence 45, Appli
11	26	0.8	1380	2 US-08-467-559B-1	Sequence 1, Appli
12	26	0.8	2406	1 US-08-396-479B-7	Sequence 7, Appli
13	26	0.8	2406	1 US-08-818-823-7	Sequence 7, Appli
14	26	0.8	19736	4 US-09-740-035-3	Sequence 3, Appli
15	25	0.7	337	4 US-09-328-111-586	Sequence 586, App
16	25	0.7	342	4 US-09-385-982-342	Sequence 342, App
17	25	0.7	523	2 US-08-628-413-1	Sequence 1, Appli
18	25	0.7	547	1 US-08-131-365B-41	Sequence 41, Appli
19	25	0.7	547	2 US-08-668-123-41	Sequence 41, Appli
20	25	0.7	567	4 US-09-385-982-427	Sequence 427, App
21	25	0.7	587	1 US-08-313-681A-3	Sequence 3, Appli
22	25	0.7	593	3 US-09-322-911-3	Sequence 3, Appli
23	25	0.7	607	4 US-09-385-982-262	Sequence 262, App
24	25	0.7	607	4 US-09-385-982-288	Sequence 288, App
25	25	0.7	629	4 US-09-385-982-389	Sequence 389, App
26	25	0.7	730	4 US-09-352-990-3	Sequence 3, Appli
27	25	0.7	770	3 US-08-865-297-5	Sequence 5, Appli

28 25 0.7 777 4 US-09-385-982-531 Sequence 531, Appli
 c 29 25 0.7 809 1 US-08-441-629-1 Sequence 1, Appli
 c 30 25 0.7 809 3 US-08-776-207-1 Sequence 1, Appli
 c 31 25 0.7 809 5 PCT-US95-09172-1 Sequence 1, Appli
 c 32 25 0.7 1225 1 US-08-286-020-1 Sequence 1, Appli
 c 33 25 0.7 1225 1 US-08-603-919-1 Sequence 1, Appli
 c 34 25 0.7 1227 1 US-08-080-386-1 Sequence 1, Appli
 c 35 25 0.7 1227 2 US-08-390-000A-1 Sequence 1, Appli
 c 36 25 0.7 1621 4 US-09-020-956-107 Sequence 107, App
 c 37 25 0.7 1621 4 US-09-030-607-107 Sequence 107, App
 c 38 25 0.7 1621 4 US-09-439-313-107 Sequence 107, App
 c 39 25 0.7 1683 4 US-09-347-803-11 Sequence 11, Appli
 c 40 25 0.7 1690 4 US-09-071-224-1 Sequence 1, Appli
 c 41 25 0.7 1813 4 US-09-071-224-3 Sequence 3, Appli
 c 42 25 0.7 1889 3 US-09-187-050-1 Sequence 1, Appli
 c 43 25 0.7 1921 2 US-08-557-128-11 Sequence 11, Appli
 c 44 25 0.7 1964 1 US-08-132-168A-31 Sequence 31, Appli
 c 45 25 0.7 1968 2 US-08-937-540-3 Sequence 3, Appli

ALIGNMENTS

RESULT 1
 US-08-072-610-1
 ; Sequence 1, Application US/08072610
 ; Patent No. 5532133
 ; GENERAL INFORMATION:
 ; APPLICANT: Barnwell, John
 ; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
 ; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Darby and Darby
 ; STREET: 805 Third Ave.
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: USA
 ; ZIP: 10022-7513
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/072,610
 ; FILING DATE: 19930602
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogoris, Adda
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 5986/07686
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212)527-7700
 ; TELEFAX: (212)753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3337 base pairs
 ; TYPE: NUCLEIC ACID
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Plasmodium vivax
 ; IMMEDIATE SOURCE:
 ; CLONE: pVMB3.3.1
 US-08-072-610-1

Query Match

100.0%; Score 3337; DB 1; Length 3337;

Best Local Similarity 100.0%; Pred. No. 0;			
Matches 3337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	gaattccggtaagtaacaactatggttcctgctatctatataaaccttactaattttatc	60
DB	1	GAATTCGGTAAAGTAACAACATATGTTTCGTATCTATATATAACCTTACTAATTTTATC	60
QY	61	tttggctttctttaaatttcattcgtctcaacagtaagataaaaaataatctataaaaactgc	120
DB	61	TTTTTGCTTTCTTTTAATTCATGCTTCAACAGTAAGATAAAAAATAATCTATAAAAACGTC	120
QY	121	tatatatacatatattcatatgaatggcatttggtaattgcgatcatcttaaatattacgta	180
DB	121	TATATATACATATATATTCATAGTGGCATTTGTGAATTTCCGATCATTTTAAATTTACGTA	180
QY	181	aaacaataatgaaaaaatt	240
DB	181	AAACAATAATTGAAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTCTACAGACGATTTAG	240
QY	241	aattggaaaatgctctgatgattgttagagtgtagatccttcaaacgcggtttag	300
DB	241	AATTGCAAAATGCTCTGATGATGTTGTAGAGTGGAGGATCCTTCAACACGACGGTTTAG	300
QY	301	aattagaagagaaaattttgatgagaattcaggtgatgatgaactcttttagatgcta	360
DB	301	AATTAGAAGAGGAAAAATTTTGATGAGAAATTCAGGTGATGATGAAACTCTTTTAGATGCTA	360
QY	361	ccccgaagatgacttgccttaacagatttccaaatgaacacgatgaggaagcaacg	420
DB	361	CCCCCAAGATGACTTTGGCTTTAACAGATTTGCCAATTTGAACACGATGAAGAACTCAACG	420
QY	421	aaacgttagatgagtggaatcattagggagaggtttccactgaagatatggaaacagaag	480
DB	421	AAACGTTAGATGAGGTGAATCATTTAGGAGAGGTTTCCACTCAAGATATGGAACAGAAAG	480
QY	481	atggttcaacagatatacggaacagaagaaggactacctggtgatattggaagagaag	540
DB	481	ATGGCTCAACAGATGATACGGAACACAGAAGAAGGACTACTCTGTGATATGGAAGGAGAAG	540
QY	541	aaagaactggcgatattggaagcaggggaagagctgggtgatitggaacgaggggaagaaa	600
DB	541	AAGAAGCTGGCGATATGGAAGCAGGGGAAGAGAGCTGGTCATTTGGNAGCAGGGGAAGAAA	600
QY	601	ctggcgatttgaagcaggggaagaaactggcgatttggaaagcaggggaagagctggcg	660
DB	601	CTGGCGATTTGGAAGCAGGGGAAGAACTGGCGATTTGGAAGCAGGGGAAGAAAGCTGGTG	660
QY	661	atttgaagcaggggaagaaactggcgatttggaaagcaggggaagagctggcgatgagcg	720
DB	661	ATTTGGAAGCAGGGGAAGAACTGGCGATTTGGAAGCAGGGGAAGAACTGGAGATGCGG	720
QY	721	aaactgaagaagagcaactggagatgcggaactgaaatggagcaactgtgtatgtag	780
DB	721	AAACTGAAGAAGAGCAACTGGAGATGCGGAACCTGAATAATGGACCACTGTGTATGTAG	780
QY	781	acacagaagatagttcagctgatgagcagaaaaagatcatattctctgtctcaagaaaaatg	840
DB	781	ACACAGAAGATAGTTCAGCTGATGGAGCAGAAAAAGTACATGTCTCTCAAGAAAAATG	840
QY	841	tacaactcccatagatgacccctcttggaaagtatttggataaagataaattt	900
DB	841	TACAACCTCCGATAGTATGATGCCCTCTTTGGAAAGTATTTTGGATAAAGATATAATTT	900
QY	901	ttgatcatataaagatttcagaccactatctgaacaaatttggcggtactgcttaaac	960
DB	901	TTGATCATATTAAGATTTTCAGGCCACTATTTCGAACAAATTTGGCGGGTACTGTCTAAAC	960
QY	961	atgttcaggagcaagaatttgcgaatgaaccctgtaccattaccagtggcagaagaccgcg	1020
DB	961	ATGTTACGGGACAGAATTTGCCAATGAACCTGTACCATTTACACGTGGCAGAGAGGCCCG	1020
QY	1021	cqcaagtaccacgggaagattagatgccactccagagatgacttcgattagatgtta	1080
DB	1021	TT	1080

DB	1021	CCCAAGTACCACGGGAAGAAATTTAGATGCCACTCCAGAGGATGACTTCGATTTAGATGTTA	1080
QY	1081	cagaatctcccagaggaagtagaatttagtattagatgaagagcaactgaagaagaatacaa	1140
DB	1081	CAGAACTCTCCGAGGAGTAGAATTTAGTATTAGATGAAGAGCAACTGAAGAAGTAATCAA	1140
QY	1141	cgaagttgggaccacggaagaaggaacccaaacccgaagaatttagatgccactccagaggtg	1200
DB	1141	CGAAGTGGGACCAACGGAAGAACCAACCGAAGAATTTAGATGCCACTCCAGAGGATG	1200
QY	1201	gatttcgcattagacaagaacttcagaaggaagaacagaagaacatagaggagaagaaa	1260
DB	1201	GATTTTCGATTTAGACAAACTGCAGAAGGAGAAACAGAAACCTAGAGGGAGAAGAAA	1260
QY	1261	cagaagaagctcagaagaagagatcagagaagaactccagaaggaagaagaagattag	1320
DB	1261	CAGAAGAAGCTGCAGAAGGAGAAGTATCAGAACAACCTCCAGAAGGAGAGAAGATTAG	1320
QY	1321	aggcaactccagaggatgatttcgcatttagatggaactacattagaagaacccgaagaaa	1380
DB	1321	AGGCAACTCCAGAGGATGATTTTCGCAATTAGATGGAACCTACATTAGAAGAAACCGAAGAA	1380
QY	1381	ctgcagaaggaagaagaacccgtagagggaagaacacgtagaggagaagaacacgtag	1440
DB	1381	CTGCAGAAGGAGAAACCCGTAGAGGGAGAAACCCGTAGAGGAGAAACCCGTAG	1440
QY	1441	egggaagaagctcagaagaagagagagaggttagaggaactccagaggtgactctcc	1500
DB	1441	AGGGAGAAGAAGCTCGACNAGGAGAAGAGAGTTAGAGGCAACTCCAGAGGATGACTTCC	1500
QY	1501	aattagaagaaccatcagggagaaggaagggaagggaagggaagggaagggaag	1560
DB	1501	AATTAGAAGAACCATCAGGAGAAGGAGAAGGGGAAGAGAGAGGAGGAGGAGAGAG	1560
QY	1561	gagaagcgttagtagcagtagccagtagtggccgaacccgtagaactagtagctctgctc	1620
DB	1561	GAGAAGCCTGTAGTAGCAGTGGCAGTAGTGGCCGAACCCGTAGAAGTAGTACTCTCTGCTC	1620
QY	1621	agcctgtcaaaccaatggtcgctccaaacgagatgaaactttatctgtatctctag	1680
DB	1621	AGCCTGTCAAAACCANTGGTCGCTCCCAACGGCAGATGAACTTTATTCGTGATATCTTAG	1680
QY	1681	ataacgatttaacgtatgcagacattacatcctttgagccattttaaaacaaatctctca	1740
DB	1681	ATAACGATTTAACGTTATGCAGACATTACATCTCTTTGAGCCATTATTATAACAATCTCTCA	1740
QY	1741	aggtatcctgtagcagagagagctgttaacagttaccatcaagaagacacctgtacaagtac	1800
DB	1741	AGGATCTCTGATGCAGAGAGGCTGTAAACAGTACCATTCAAGAGACACCTGTACAAGTAC	1800
QY	1801	cagtgagcagtagggcccgccaagaagtgcgaacggaagaatttgatgcaactcccaagagg	1860
DB	1801	CAGTGGCAGTAGGGCCCGCCGAAGAAGTGCCAAACGGAAGAATTGATGCAACTCCAAGAGG	1860
QY	1861	acgatttcgaattagaagaactgcagaagctccagagaaggaagaatttagtattagaag	1920
DB	1861	ACGATTTCGAATTAGAAGSACTGCAGAAGCTCCAGAGGAGGAGAAATTAGTATTAGAAAG	1920
QY	1921	gagaagagaaccaaaggaagaagccaagaagagagccaagaagaagaagaagtgc	1980
DB	1921	GAGAAGGAGAACCAACGGAAGAGAGCCACAGAGAGGAGAGCCACAGAGGAGAGTGC	1980
QY	1981	cagaagaagaatttagaggcaactccagaggacgatttcgaattagaagaaccaaagagg	2040
DB	1981	CAGAAGAAGAAATTAGAGGCAACTCCAGAGGACGATTTTCAATTAGAAGAAACCAACAGAGG	2040
QY	2041	aagaagttagaagaacccgtagagggcgaagaacatgcagaaggaagaagaagtggagaag	2100
DB	2041	AAGAAGTGAAGAACCCTTAGAGGGCGAAGAAACTGCAGAAGGAGAAAGTGGGAAGAGG	2100
QY	2101	tacctgcagaagtagaagaagtgggaaggttaccttcgagaagtagaagaagtgggaagag	2160
DB	2101	TACCTGCAGAAGTGAAGAAGTGAAGAGGTACCTGCAGAAGTGAAGAGTGAAGAGG	2160

QY 361 cccccgaagatgactttgcttaacagatttgccaattgaaacgcatgaggagtcacacg 420
|||||
Db 361 CCCCCGAAGATGACTTTGCTCTTAACAGATTGCGCAATTGAACAGCATGAGGAAGTCAACG 420

QY 421 aaacgttagatgagtgatcatattagagagaggtttccactgaagatatggaaacagaag 480
|||||
Db 421 AAACGTTAGATGGAGTGAATCAATTAGGAGAGGTTTCCACTCAAGATATGGAACAGAAAG 480

QY 481 atgctcaacagatgatacggaacagaagaagactacctggtgatatggaggagaag 540
|||||
Db 481 ATGGCTCAACAGATGATACGGAACAGAGAGGACTACTGCTGATATGGAAGGAGAG 540

QY 541 aagaagctggcgatatggaagcagggggaagaagctggtgatttggaagcagggggaagaaa 600
|||||
Db 541 AAGAAGCTGGCGATATGGAAGCAGGGGAAGAGCTGGTGATTGGAAGCAGGGGAAGAAA 600

QY 601 ctggcatttgaagcaggggaagaacitggcatttgaagcagggggaagactggtg 660
|||||
Db 601 CTGGCATTGGAAGCAGGGGAAGAACCTGGCGATTGGAAGCAGGGGAAGAGCTGGTG 660

QY 661 atttgaagcagggggaagaactggcatttgaagcagggggaagaactggagatgcgg 720
|||||
Db 661 ATTTGGAAGCAGGGGAAGAACTGGCGATTGGAACAGGGGAAGAACTGGAGATGCCG 720

QY 721 aaactgaagaagagcaactggagatgcggaaactgaaatggagcaactggtatgtag 780
|||||
Db 721 AAACCTGAAGAAGAGCAACTGGAGATGCGGAACCTGAAATGGCAACTGCTATGTAG 780

QY 781 acacagaataagtcagctgatggagcagaaaaagtcacattctctctcaagaataag 840
|||||
Db 781 ACACAGAATAGTTCAGCTGATGGAGCAGAAAAGATACATGTTCTGCTCAAGAAAATG 840

QY 841 tacaacctgcgcatagtaagtgcctctcttggagagttttggataaagataaatt 900
|||||
Db 841 TACAACCTGCGCATAGTAATGATGCCCTCTTTGGAAGTATTTGGATAAAGATAATATTT 900

QY 901 ttgatcatattaaagatttcagccactattcgaaacaaatttggcgggtactgctaaac 960
|||||
Db 901 TTGATCATATTAAAGATTTTCGAGCCCACTATTGCAACAAATTTGGCGGGTACTGCTAAAC 960

QY 961 atgttcgggacaagaattgccaatgaacacctgtaccattaccagtggcagaagaccg 1020
|||||
Db 961 ATGTTACGGGACAAGAAATGCCCAATGAACCTGTACCATTAACAGTGGCAGAGACCCG 1020

QY 1021 cgcaatccagcgggaagaattagatgcccactccagagatgacttcgattagatgtta 1080
|||||
Db 1021 CGCAAGTACCAGCGGAAGAATTAGATGCCACTCCAGAGGATGACTTCGCATTAGATGTTA 1080

QY 1081 cagaactcccgaggagtagaattagattagatgaagagcaactgaagaagaatcaa 1140
|||||
Db 1081 CAGAATCTCCGAGGAAGTAGAATTAGTATTAGATGAAGAGGCAACTGAAGAAGAAATCAA 1140

QY 1141 cggaagtggaccacaggaagagccaccgaagaatttagatgccactccagagagtg 1200
|||||
Db 1141 CGGAAGTGGGACCACCGGAAGAGGACCAACCCGAAATTAGATGCCACTCCAGAGGATG 1200

QY 1201 gatttcgattacacgaactcgagaagcagaagaacacgtagaggagagaagaaa 1260
|||||
Db 1201 GATTTCCGATTACACGAAACTGCAGAAGGAGAAACAGAAACGTFAGAGGGGAAGAAA 1260

QY 1261 cagaagaagctcgagaagagagtagatcagaagaactccagaagagagaagaagtag 1320
|||||
Db 1261 CAGAAGAAGCTGCGAAGGAGAGTAGTATCAGAAGAACTCCAGAAGGAGGAAGAGATTAG 1320

QY 1321 aggcaactccagagtagtatttcgattagatggaactacattagaagaaaccgaagaaa 1380
|||||
Db 1321 AGGCAACTCCAGAGGATGATTTCCCAATTAGATGGAACCTACATTAGAAGAAACCGAAGAAA 1380

QY 1381 ctgcagaagagagaacacgtagaggagagaacacgtagaggagagaagaacgtag 1440
|||||
Db 1381 CTCGAGAAGGAGAGAAACCGTAGAGGGAGAGAAACCGTAGAGGGGAGAGAACCGTAG 1440

QY 1441 agggagaagaagctcgagaagagagaagattagagcaactccagaggatgacttcc 1500
|||||
Db 1441 AGGGAGAAGAAGCTCGAAGGAGAAGAGAGTTAGAGCAACTCCAGAGGATGACTTCC 1500

QY 1501 aattagaagaaccatcagagaagagaaggggaagagaaggaaggaagagaag 1560
|||||
Db 1501 AATTAGAAGAACCATCAGGAGAAGGAGGGAAGGAGAGGGAAGGGAAGGAGAGAG 1560

QY 1561 gagaagcgttagtagcagtgccagtagtggccgaaccggtagaagtagtactcctgtcc 1620
|||||
Db 1561 GAGAAGCGTTAGTAGCAGTGCCAGTAGTGGCCGAACCGGTAGAGTAGTACTCTGCTC 1620

QY 1621 agcctgtcaaaccaatgctgcctccaaaggcagatgaaactttatctgtgatattag 1680
|||||
Db 1621 AGCCTGTCAAAACCAATGCTGCTCCAACGGCAGATGAAACTTTATTCGTTGATATCTTAG 1680

QY 1681 ataacgatttaacgtatcagacattacatccttggaccattatttaacaaatccctca 1740
|||||
Db 1681 ATAACGATTTAACGTATGCGAGACATTACATCTCTTTGAGCCATTATTTAAACAAATCTCTCA 1740

QY 1741 aggatcctgatgcaggagaggtgctaaacagttaccatcaaaaggaagcactgtacaagtac 1800
|||||
Db 1741 AGGATCCTGATGCAGGAGAGGCTGTAACAGTAGTACCATCAAAAGGAAGCACCTGTACAAGTAC 1800

QY 1801 cagtggcagtagggcccgcgaagaagtgcacaacggaagaattgatgcaactccaaagg 1860
|||||
Db 1801 CAGTGGCAGTAGGGCCCGCGCAAGAAGTGCCAAACGGAAGAATTGATGCAACTCCAAGAG 1860

QY 1861 acgatttcgaattagaagaactgcagaagctccagaggaaggaattagattagaag 1920
|||||
Db 1861 ACGATTTGCAATTAGAGGAACCTGCAGAACTCCAGAACTCCAGAGGAAGAGANTAGTATTAGAAG 1920

QY 1921 gagaagggaacccaacggaagagcgaagaggaagagccacaagaagaggaagtgc 1980
|||||
Db 1921 GAGAAGGAGAACCAACGGAAGAAGAGCCAAAGAAAGAGGAGAGCAACAGAAAGGAGAAGTGC 1980

QY 1981 cagaagaagaattagaggcaactccagaggacgatttcgaattagaagaacccaacaggag 2040
|||||
Db 1981 CAGAAGAAGAAATTAGAGGCCAACTCCAGAGGACGATTTCGAATTAGAAGAACCACAGCAGGAG 2040

QY 2041 aagaagtagaagaacccgttagaggcgcaagaactgcagaaggaagaagtggaagag 2100
|||||
Db 2041 AAGAAGTAGAAGAAACCGTtagaggcgcaagaactgcagaaggaagaagtggaagag 2100

QY 2101 taccctgcgaagatagaagaagtgggaaggttaccctgcgaagtagaagaagtgggaagag 2160
|||||
Db 2101 TACCTGCGAAGTAGAAGAAGTGGGAAGGTACCTGCGAAGTAGAAGAAGTGGGAAGAGG 2160

QY 2161 taccagaagaagttagaagaggttaccgcgagaagttagaagaagtgggaaggttaccagaag 2220
|||||
Db 2161 TACCAGAAGAAGTAGAAGAGGTACCCGCGAAGTGAAGAGTGAAGAGGTGGAAGAGTACCAGAAG 2220

QY 2221 aagtggagaaggttaccagaagaagtgggaaggttaccagaagaagtgggaagaggtlaccag 2280
|||||
Db 2221 AAGTGGGAAGAGGTACCAGAAGAAGTGGGAAGGTACCAGAAGAAGTGGGAAGAGGTACCAG 2280

QY 2281 aagaagtggaagaagtgggaagtagaagaagttagaggttaccagcgttagtagaagtag 2340
|||||
Db 2281 AAGAAGTGGGAAGAAGTGGGAAGAAGTAGAAGAAGTGAAGGTACAGCGGTAGTAGAAGTAG 2340

QY 2341 aagtaccagcgttagtagaagaaggttccagaagaagttagaagaagaaggaagagag 2400
|||||
Db 2341 AAGTACCAGCGGTAGTAGAAGAAGAGGTGCCAGAAGAGTAGAAGAAGGAAGGAAGAGG 2400

QY 2401 aagaaccagtagaggaagaagatgtattacaattagtaataccatcggaagaagataac 2460
|||||
Db 2401 AAGAACCCAGTAGAGGAGAGAGATGTTATCAATTAGTATACCATCGGAAGAGATATAC 2460

QY 2461 aatagacaaaccaaagaagacgaatttaggctctggaaatttattatcatcatgcagatgc 2520
|||||
Db 2461 AATTAGACAAACCAAAAGAAAGACGAATTAGGCTCTGGAATTTTATCTATCATCGCATGC 2520

QY 2521 actaccaagacgttccaaagggaatttatggaagaagaagaagaagagtagtcatccat 2580

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 96243720
FILING DATE: 13-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: KILYK JR., JOHN
REGISTRATION NUMBER: 30763
REFERENCE/DOCKET NUMBER: 72882
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 616-5600
TELEFAX: (312) 616-5700
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 90 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE: 5'-end is phosphorylated
US-09-065-058-16

Query Match 0.8%; Score 28; DB 4; Length 90;
Best Local Similarity 100.0%; Pred. No. 0.021; 0; Indels 0; Gaps 0;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 200 ttttttttttttttttttttttttctta 227
|||||
Db 49 TTTTtttttttttttttttttttttttCTTA 76

RESULT 6
US-08-828-007-7
Sequence 7, Application US/08828007
Patent No. 5866412
GENERAL INFORMATION:
APPLICANT: Chen, Hong
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
THE DIAGNOSIS AND TREATMENT OF NEUROPSYCHIATRIC
DISORDERS
TITLE OF INVENTION: DISORDERS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036/2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,007
FILING DATE: 27-MAR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Coruzzi, Laura A
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7853-069
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-790-9090
TELEFAX: 212-869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs

TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: Antisense oligonucleotide
LOCATION: 1...50
OTHER INFORMATION:
US-08-828-007-7
Query Match 0.8%; Score 27; DB 2; Length 50;
Best Local Similarity 100.0%; Pred. No. 0.054; 0; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 aaaaattttttttttttttttttttt 221
|||||
Db 24 AAAAatttttttttttttttttttttT 50
RESULT 7
US-07-867-106-4
Sequence 4, Application US/07867106
Patent No. 5389526
GENERAL INFORMATION:
APPLICANT: Slade, Martin B
APPLICANT: Chang, Andy C M
APPLICANT: Williams, Keith L
TITLE OF INVENTION: Improved plasmid vectors for Cellular
TITLE OF INVENTION: Slime Moulds of the Genus Dictyostelium
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock Washburn Kurtz Mackiewicz & No. 5389526ris
STREET: One Liberty Place 46th Floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/867,106
FILING DATE: 19920625
PRIOR APPLICATION DATA:
APPLICATION NUMBER: AU PJ 7187
APPLICATION NUMBER: PCT/AU90/00530
FILING DATE: 02-NOV-1989
ATTORNEY/AGENT INFORMATION:
NAME: Feeney, Joanne Longo
REGISTRATION NUMBER: 35,134
REFERENCE/DOCKET NUMBER: RICE-0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-568-3100
TELEFAX: 215-568-3439
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 3138 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
ANTI-SENSE: NO
US-07-867-106-4
Query Match 0.8%; Score 27; DB 1; Length 3138;
Best Local Similarity 100.0%; Pred. No. 0.044; 0; Indels 0; Gaps 0;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 195 aaaaattttttttttttttttttttt 221

Db 1938 AAAAAATTTTTTTTTTTTTTTTTTTTTTTT 1964
|||||

RESULT 8
US-09-244-794A-1/c
; Sequence 1, Application US/09244794A
; Patent No. 6214553
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350006
; CURRENT APPLICATION NUMBER: US/09/244,794A
; CURRENT FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: 60/035,963
; PRIOR FILING DATE: 1997-01-27
; PRIOR APPLICATION NUMBER: 60/064,491
; PRIOR FILING DATE: 1997-11-06
; PRIOR APPLICATION NUMBER: 09/007,005
; PRIOR FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 76
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
US-09-244-794A-1

Query Match 0.8%; Score 26; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttc 225
|||||

Db 69 TTTT

RESULT 9
US-09-247-190-1/c
; Sequence 1, Application US/09247190
; Patent No. 6261804
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350005
; CURRENT APPLICATION NUMBER: US/09/247,190
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-21
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 76
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
US-09-247-190-1

Query Match 0.8%; Score 26; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttc 225
|||||

Db 69 TTTT

Query Match 0.8%; Score 26; DB 4; Length 76;
Best Local Similarity 100.0%; Pred. No. 0.13;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttc 225
|||||

Db 69 TTTT

RESULT 10
US-08-150-331-45/c
; Sequence 45, Application US/08150331
; Patent No. 5516512
; GENERAL INFORMATION:
; APPLICANT: DORSERS J., LAMBERTUS C.
; APPLICANT: VAN LEEN, ROBERT W.
; TITLE OF INVENTION: MUTANTS OF HUMAN INTERLEUKIN-3
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/150,331
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/651,437
; FILING DATE: 05-FEB-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: GRACEY, NANCY J.
; REGISTRATION NUMBER: 28,216
; REFERENCE/DOCKET NUMBER: 24615-20010.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 596 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..435
US-08-150-331-45

Query Match 0.8%; Score 26; DB 1; Length 596;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttc 225
|||||

Db 561 TTTT

RESULT 11
US-08-467-559B-1
; Sequence 1, Application US/08467559B
; Patent No. 5928890
; GENERAL INFORMATION:
; APPLICANT: LI, YI
; TITLE OF INVENTION: HUMAN AMINE RECEPTOR

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN AND FOX, P.L.L.C.
STREET: 1100 NEW YORK AVENUE, NW, SUITE 600
CITY: WASHINGTON
STATE: DC
COUNTRY: UNITED STATES OF AMERICA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/467,559B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: STEFFE, ERIC K
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.084.0000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1380 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 252..1262
US-08-467-559B-1

Query Match 0.8%; Score 26; DB 2; Length 1380;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 196 aaaaattttttttttttttttttttttt 221
Db 158 AAAATTTTTTTTTTTTTTTTTTTTTT 183

RESULT 12
US-08-396-479B-7/C
Sequence 7, Application US/08396479B
Patent No. 5612455
GENERAL INFORMATION:
APPLICANT: HOEY, Timothy
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/396,479B
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59450-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 211..2337
US-08-396-479B-7

Query Match 0.8%; Score 26; DB 1; Length 2406;
Best Local Similarity 100.0%; Pred. No. 0.11;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 200 tttttttttttttttttttttttttttttc 225
Db 2398 TTTTTTTTTTTTTTTTTTTTTTTC 2373

RESULT 13
US-08-818-823-7/C
Sequence 7, Application US/08818823
Patent No. 5708158
GENERAL INFORMATION:
APPLICANT: HOEY, Timothy
TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY
NUMBER OF SEQUENCES: 18
CORRESPONDENCE ADDRESS:
ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,823
FILING DATE: 14-MAR-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/396,479
FILING DATE: 02-MAR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Osman, Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: A-59450-1/RAO
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 210 277299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2406 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 211..2337
US-08-818-823-7

Seq. Loc. Similarity 100.00; P-Id. No. 0.11,
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

```

RESULT 16
US-09-385-982-342
; Sequence 342, Application us/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; TITLE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 342
; LENGTH: 342
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(342)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-342

```

```

;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: misc_feature
;
; LOCATION: (1)...(342)

```

; LOCATION: (1)...(342)
 ; OTHER INFORMATION: n = A,T,C or G
 US-09-385-982-342

```

Query Match          0.7%; Score 25; DB 4; Length 342;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels

Qy    200  ttttttttttttttttttttttgg 224
          |||||
Dd    9    ttttttttttttttttttttttgg 33

RESULT 17
US-08-628-413-1/c
; Sequence 1, Application US/08628413
; Patent No. 5834192
; GENERAL INFORMATION:
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Murry, Lynn E.
; TITLE OF INVENTION: HUMAN CACHEXIA ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,413
; FILING DATE: Filed Herewith

```

```

; SOFTWARE: FASTSEQ VERSION 1.3
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/628,413
; FILING DATE: Filed Herewith
;

```

```
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0062 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 523 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; IMMEDIATE SOURCE:
; LIBRARY: BRSTTUT01
; CLONE: 607227
; US-08-628-413-1

Query Match          0.7%; Score 25; DB 2; Length 523;
Best Local Similarity 100.0%; Pred. No. 0.31;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 197 aaattttttttttttttttttttttttt 221
| | | | | | | | | | | | | | | | | |
DB 477 AAATTTTTTTTTTTTTTTTTTTTTTTT 453

RESULT 19
US-08-131-365B-41/c
; Sequence 41, Application US/08131365B
; Patent No. 5527690
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/131,365B
; FILING DATE: 01-OCT-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
```

```
; DESCRIPTION: /desc = "DNA"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..111
; US-08-131-365B-41

Query Match          0.7%; Score 25; DB 1; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttgtt 224
| | | | | | | | | | | | | | | | | |
DB 540 TTTTTTTTTTTTTTTTTTTTTTGT 516

RESULT 19
US-08-668-123-41/c
; Sequence 41, Application US/08668123
; Patent No. 5891631
; GENERAL INFORMATION:
; APPLICANT: Brown, Michael S.
; APPLICANT: Briggs, Michael R.
; APPLICANT: Wang, Xiaodong
; APPLICANT: Goldstein, Joseph L.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS RELATING
; TO STEROL REGULATORY ELEMENT BINDING
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 64
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: U.S.A.
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/668,123
; FILING DATE: 14-JUN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/131,365
; FILING DATE: 01-OCT-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: UTSD:372/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 547 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..111
; US-08-668-123-41

Query Match          0.7%; Score 25; DB 2; Length 547;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 200 tttttttttttttttttttttttttttttttgtt 224
|||||
Db 540 TTTT

RESULT 20
US-09-385-982-427
; Sequence 427, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE REFERENCE: CCNDA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-08-30
; EARLIER FILING DATE: 1999-06-08
; EARLIER FILING DATE: 1999-01-27
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 427
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(567)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-427

Query Match 0.7%; Score 25; DB 4; Length 567;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 tttttttttttttttttttttttttttttttgtt 224
|||||
Db 14 tttttttttttttttttttttttttttttttgtt 38

RESULT 21
US-08-313-681A-3/C
; Sequence 3, Application US/08313681A
; Patent No. 5618675
; GENERAL INFORMATION:
; APPLICANT: Larrick, James W.
; APPLICANT: Wright, Susan C.
; APPLICANT: Hirata, Mishimasa
; TITLE OF INVENTION: Human Cationic Proteins Having
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: One Market Plaza, Steuart Tower, Suite 2000
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/08/313,681A
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.

; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 15325-9-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-313-681A-3

Query Match 0.7%; Score 25; DB 1; Length 587;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 200 tttttttttttttttttttttttttttttttgtt 224
|||||
Db 585 TTTT

RESULT 22
US-09-322-911-3/C
; Sequence 3, Application US/09322911
; Patent No. 6103888
; GENERAL INFORMATION:
; APPLICANT: Larrick, James W.
; APPLICANT: Wright, Susan C.
; APPLICANT: Hirata, Mishimasa
; APPLICANT: Balint, Robert F.
; TITLE OF INVENTION: Human Cationic Proteins Having
; TITLE OF INVENTION: Lipopolysaccharide Binding and Anti-Coagulant Activity
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/322,911
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/691,280
; FILING DATE: August 1, 1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/916,761
; FILING DATE: July 17, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/916,765
; FILING DATE: July 17, 1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/06731
; FILING DATE: July 15, 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/313,681
; FILING DATE: September 27, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 15325-000920
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400

TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 587 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-322-911-3

Query Match 0.7%; Score 25; DB 3; Length 587;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 585 TTTTtttttttttttttttttttttttttttt 561

RESULT 23
US-09-385-982-262
; Sequence 262, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 262
; LENGTH: 593
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(593)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-262

Query Match 0.7%; Score 25; DB 4; Length 593;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 6 ttttttttttttttttttttttttttttttt 30

RESULT 24
US-09-385-982-288
; Sequence 288, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393

; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 288
; LENGTH: 607
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(607)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-288

Query Match 0.7%; Score 25; DB 4; Length 607;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 6 ttttttttttttttttttttttttttttttt 30

RESULT 25
US-09-385-982-389
; Sequence 389, Application US/09385982
; Patent No. 6262334
; GENERAL INFORMATION:
; APPLICANT: ENDEGE, WILSON O., ET AL.
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION
; FILE OF INVENTION: PRODUCTS: II
; FILE REFERENCE: CCDNA-260XX
; CURRENT APPLICATION NUMBER: US/09/385,982
; CURRENT FILING DATE: 1999-08-30
; EARLIER APPLICATION NUMBER: 09/328,111
; EARLIER FILING DATE: 1999-06-08
; EARLIER APPLICATION NUMBER: 60/117,393
; EARLIER FILING DATE: 1999-01-27
; EARLIER APPLICATION NUMBER: 60/098,639
; EARLIER FILING DATE: 1998-08-31
; NUMBER OF SEQ ID NOS: 544
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 389
; LENGTH: 629
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(629)
; OTHER INFORMATION: n = A,T,C or G
US-09-385-982-389

Query Match 0.7%; Score 25; DB 4; Length 629;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 200 ttttttttttttttttttttttttttttttt 224
|||||
Db 5 ttttttttttttttttttttttttttttttt 29

RESULT 26
US-09-352-990-3/c
; Sequence 3, Application US/09352990
; Patent No. 6255090
; GENERAL INFORMATION:
; APPLICANT: Famodu, Layo O.
; APPLICANT: Orozco, Buddy
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Plant Aminoacyl-tRNA Synthetase

Query Match 0.7%; Score 25; DB 3; Length 770;

```

RESULT 29
US-08-441-629-1/c
; Sequence 1, Application US/08441629
; Patent No. 5766923
; GENERAL INFORMATION:
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Kinoshita, No. 5766923Iyuki
; TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441.629
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

```

Search completed: June 14, 2002, 20:29:19
Job time: 12413 sec